



AAPOR

AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH

May 18-21, 2017

Sheraton New Orleans
New Orleans, Louisiana

72nd
Annual
Conference

Abstract Book

**Embracing Change and
Diversity in Public Opinion
and Social Research**

www.aapor.org/conference



#AAPOR

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Eye-Tracking Methodology: What Affects the Processing of Agree/Disagree and Item-Specific Questions?

Jan Karem Höhne *University of Göttingen*

Timo Lenzner *GESIS - Leibniz Institute for the Social Sciences*

Since Likert's (1932) well-known article "A Technique for the Measurement of Attitudes", the use of agree/disagree (A/D) questions has become increasingly popular in empirical social research, because it appears to be possible to measure several constructs with a single response format. Fowler (1995), by contrast, argues that using item-specific (IS) questions represents a simpler, more direct, and more informative method than using agree/disagree questions. Until now, this assumption has never been tested empirically. In the current study, we investigate cognitive processing of A/D and IS questions in web surveys using eye-tracking methodology. Recording respondents' eye movements, we are able to investigate how respondents process the survey questions and response categories to determine their information processing. We conducted an eye-tracking experiment with two groups: the first group (n = 44) received three A/D single questions dealing with different political issues (agree/disagree condition). The second group (n = 40) received three IS counterparts dealing with the same question content (item-specific condition). The results of our investigations indicate that IS questions encourage a deeper cognitive processing than A/D questions. Interestingly, the eye-tracking data reveal that this fact is only observable for the processing of the response categories. There are no differences with respect to the question stems; indicating that "question comprehension" seems to be equal. We therefore argue that the observed differences between these two question formats are directly attributed to a more active and more intensive cognitive processing of the IS response categories.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Impact of Mixed-Mode Recruitment and Data Collection on Sample Representativeness and Survey Estimates for a Probability-Based Household Panel

Dan Malato *NORC at the University of Chicago*

David Sterrett *NORC at The University of Chicago*

Jennifer Benz *NORC at the University of Chicago*

Trevor Thompson *NORC at the University of Chicago*

J. Dennis *NORC*

Vicki Pineau *NORC at the Univ of Chicago*

Nadarajasundaram Ganesh *NORC at the University of Chicago*

Most household panels in the US recruit their samples using the online mode and then also conduct their surveys online. The modes employed for panel recruitment and data collection are the same. In addition to relying on non-probability sampling, such mono-mode household web panels theoretically have at least two shortcomings with respect to sample coverage. First, regarding the mode of recruitment to the panel itself, web-only panels under-represent the fraction of the US population that does not have internet access and under-represents infrequent web users. Second, regarding the mode of data collection for panel-based surveys, web-only panels under-represent the fraction of the population that would prefer to participate in actual surveys by non-web modes, such as by telephone. Using data from NORC's AmeriSpeak Panel surveys (which is based on mixed-mode recruitment and data collection), the authors will document the impact (i) on sample composition from recruiting households by multi-mode methods and (ii) on survey estimates by using multi-mode for specific attitude, knowledge, and opinion surveys. Analyses demonstrate that the use of multi-mode recruitment improves the representativeness of the constituted household panel with respect to boosting sample representativeness for low-income households, Hispanic ethnicity, and seniors. Analyses also demonstrate that panel respondents completing surveys by telephone tend to have less positive views towards science and consumer technology, as well as overall lower scientific literacy. The research provides significant insights into specific ways that providing non-web modes for recruitment and data collection can enhance sample representativeness and influence survey estimates on attitudinal and opinion measures, while also identifying instances of convergence of mono-mode and multi-mode survey estimates.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Meet the Independents: Using Cluster Analysis to Construct a Typology of Independent Voters

Ayellet Pelled *University of Wisconsin - Madison*

Megan Duncan *University of Wisconsin - Madison*

Song Wang *University of Wisconsin - Statistics*

Yin Wu *University of Wisconsin - Madison*

Moonhoon Choi *University of Wisconsin - Madison*

Jiyoun Suk *University of Wisconsin - Madison*

Douglas McLeod *University of Wisconsin - Madison*

Much of the recent research on political party identification has focused on partisan voters and the consequences of political polarization. But independent voters are also an important feature of the political landscape, but far fewer studies have examined them. Too often researchers treat independent voters as a monolithic entity, using a single response category to identify them. But as made clear by the 2016 presidential election, there are vast differences among independent voters. The independent voters who latched onto the Trump movement were different from those with the Sanders movement. The independent voters who occupied the space between the Democratic and Republican parties were very different from those on the far left and far right. Some independents have beliefs that align with the core of a party, but feel that the party has abandoned the party's core principles (Walter, 2014). Additionally, the number of independent voters is on the rise, and their influence on election outcomes is growing. A recent Gallup poll showed party identification was near all-time-lows while 42 percent of Americans identified as independent (Jones, 2016). With these factors in mind, this study takes a closer look at the values and political attitudes that differentiate independent voters. Data for the analysis were collected using an online survey of 2,582 eligible voters provided by Survey Sampling International, of which 486 people identified as independents, and an additional 544 indicated they were independent but leaned toward one party. These independent voters were divided into clusters based on measures of political values and worldviews (e.g. political interests, political cynicism, political tolerance, authoritarianism, libertarianism, federalism). These clusters were further defined in terms of their demographic characteristics and linked to outcomes such as issue orientations, voting behaviors, and media use patterns.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Implications of Utilizing Consumer Grade GPS Receivers within a Transportation Research Application

Josh DeLaRosa *Abt Associates*

When paired with travel diaries, Global Positioning System (GPS) devices can provide transportation planners and modelers with rich transportation data. The GPS devices offer high resolution spatial and temporal data, which supplement diaries with attitudinal data (e.g., trip purpose). The GPS devices can also minimize the underreporting of trips in a diary by collecting data passively and continually. Common practice in transportation research is to provide respondents with consumer grade GPS devices, equivalent to the chipsets found in smartphones. However, urban canyons, weather and tree canopies can obstruct satellite reception resulting in the deterioration of GPS accuracy (type I and/or type II error). This error could ultimately lead to the introduction of error into transportation models. To overcome the deterioration of GPS accuracy, researchers can process the data and filter out suspected erroneous GPS points. To better understand possible measurement error associated with consumer grade GPS devices, this study will test and compare several GPS loggers. This study will also test different methods of processing GPS data. The results of these test may help researchers better predicting conditions leading to GPS measurement error as well as quantifying the limitations and advantages of collecting GPS data within a transportation application.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

The Practice of SMS Survey as Public Opinion Research in Japan

Kentaro Watanabe *Osaka University*

Yasuyuki Saito *The Asahi Shimbun*

In this study, we test the SMS (Short Message Service) mode and compare the result with the Asahi Shimbun CPRDD (Cell Phone RDD) data to examine the response tendencies. In Japan, the SMS mode has become available since 2011. We are aware that one market research company conducted the research, however it is not open to the public. In this research, the sample frame is Softbank mobile users. We sent the SMS message to a random-sampled 10,900 numbers and 5,724 received the message. 102 (1.8%) of them accessed the web site and 49 (0.9%) answered. For the CPRDD, 5,500 numbers were sampled and 1,132 (20.6%) answered. The accessing rate and the response rate are different between weekday and weekend. The accessing became stable after 10 hours on weekday, and the accessing rate on weekend grows similarly by the eighth hour. However, the accessing rate on the weekend increases in next 4 hours. Among the two sets of researches, the number of men was bigger than that of women and the group of young people was large enough. The result shows the similarity that the Liberal Democratic Party of Japan collected a vote of 31.13% and The Democratic Party got that of 16.85% in CPRDD, while the former is 27.66% and the latter is 19.15% in SMS. The orders for other parties are also similar. To summarize, the characteristics of SMS are (1) more people who are young and use cell phones only are included, and (2) the approval rating and the results of the election are similar to those of CPRDD. However, SMS has a problem of low response rate. As for the solution, it seems to be necessary to identify the factors which have effects on the rate.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Simulating the Effects of Changing Calling Parameters and Workload Size on Calling Efficiency: Insights from 2016 American Community Survey CATI Paradata

Gregory J. Mills *U.S. Census Bureau - Washington, DC*

The American Community Survey (ACS) has experienced downward productivity and efficiency in the Computer-Assisted Telephone Interview (CATI) mode over the past several years, as has been a common trend among many surveys that include a telephone data collection mode. The purpose of this paper was to better understand the ACS CATI operation in terms of how calling efficiency varies both across rounds of callback attempts and across the ranked workload. Using six months of 2016 CATI paradata, the analysis first shows that the proportion of the CATI workload that receives a completed interview has fallen considerably in recent years. Contrarily, the proportion of cases that are closed out due to never reaching the sample unit have substantially increased. Secondly, we found that cases where a CATI interviewer made contact in earlier rounds of calling had considerably higher completion rates than cases that did not make early contact. Finally, simulations showed that several operational changes could improve calling efficiency, including: 1) lowering the maximum number of callback attempts, 2) decreasing the size of the CATI workload, or 3) by using a combination of both methods. However, based on the data used in the simulations, both reducing the workload size or using a combination of both methods appear to provide a greater efficiency boost than by only lowering the maximum number of callback attempts.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

NYC Non Profit Community Disaster Preparedness- Response and Recovery Readiness

Janet L Streicher *Baruch College, CUNY*

Micheline Blum *Baruch College/CUNY*

Douglas Muzzio *Baruch College/CUNY*

Experience after the September 11, 2001 terrorist attacks and Superstorm Sandy (2012) in NYC demonstrated the important response and post-disaster role played by the human services non-profit community. Generally, response to such events have been unplanned and not optimally executed. The purpose of this study, conducted by Baruch College Survey Research and commissioned by the NYC Human Services Council and Department of Health and Human Hygiene, was to understand the current level of community preparedness and to make recommendations to improve readiness. 210 responded to a survey received by 582 members of the Human Services Council and their constituent agencies with a response rate of 36%. Respondents were primarily agency heads and other senior administrators. The survey, consisted of sixty four questions focused on organizational size, mission (areas served, groups of people to be assisted, types of services provided), emergency and business operations plans, communications, and coordination with government agencies. The survey was multi-mode- it was primarily online with email reminders and by telephone to maximize survey response. Findings indicate that many organizations have developed plans, have facilities and emergency supplies (e.g., blankets), and trained their staff. However, other critical operational steps to ensure plan implementation and coordination with other agencies have not been taken. Funding to support these operations has not been provided or identified.. One significant issue is that linkages have not been established between human services organizations and government agencies for leading disaster response and recovery efforts. Human Service organizations that provide services during a disaster are not generally aware of or engaged in government disaster planning. Findings and Recommendations will be provided on the poster.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Are 'Active' Landline Numbers Really Active? The Effect of Landlines not Used for Incoming Calls on True Landline Phone Penetration

James Dayton *ICF International*

Robert Tortora *ICF International*

Alex Coleo *mFour*

The U.S. household population can be divided into a number of groups based on telephone service. These include no service, landline service only, cell service only, or some combination of both landline and cell service. Most sample weighting methods only account for the categories about without any attempt break down dual-households. This group can be divided further into those who exclusively or mostly rely on their cell phones and those who rely exclusively or mostly on landlines. This presentation will explore the significance of over representation of landline phones used in dual-frame weighting schemes from households who are considered to be dual-frame but actually relay exclusively or mostly on cellular. ICF is collaborating with our mobile panel partner mFour to conduct a large scale survey of households to determine how households relay on their cell and landlines. We will profile over 15,000 new panelists to mFour's mobile platform to determine household landline service status and if calls are accepted on this landline. For those who do not use their landline for incoming calls, we will also ascertain if the household requested the service for emergencies or outbound use only or if the service is bundled as part of a cable television/internet service package. We will weight our data from households whom can be reached by landline and compare it with the estimates of landline households researchers currently use to correct dual-frame estimates. We hypnotize using this revised estimate of actual landline households whom accept calls over their landline will provide more accurate dual-frame sample corrections. In addition we will conduct a multivariate analysis of the characteristics of those households with working landline numbers that are not used to receive calls.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Phone 2.0 – Technology Democratizes Access to Voice-based Survey Methods for Better Results

Mary McDougall *Survox Inc*
Allen Porter *Survox Inc*.

Healthcare services and social science research depend on reaching and interviewing a representative audience. With online survey response rates dropping, researchers are turning again to phone — but this time using automated phone interviewing (IVR) for cost-effective respondent targeting. Learn how some of the major service providers are having success with a new generation of voice-based methods to reach seniors and non-internet users in certain socio-economic segments. Specifically we'll cover the differences in response rates for multiple collection methods and how they differ by length of survey and the sponsor's relationship to the respondent. We'll also explore new AI methods such as these. • Text-to-Speech – Personalize the recruitment and interview to enhance user experience and response rates • Speech-to-Text – Automate analysis of open-ends to quickly mine responses for deeper insights and faster decision-making and actions. The session will also cover automated phone survey best practices in questionnaire design, sample management, recording storage, and response analysis as well as overview the advances in cross-vendor integrations that enable seamless multi-mode workflows.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Improving the Net-Promoter Score: Applying Survey Methodology Insights to the 'Ultimate Question'

Sebastian Lundmark *Stanford University*

Jon Krosnick *Stanford University*

Ellen Konar *Stanford*

Matt Berent *Matt Berent Consulting*

Yphtach Lelkes *University of Pennsylvania*

Ari Malka *Yeshiva University*

Daniel Schneider *Regioconnect*

Randall Thomas *GfK Custom Research*

Ana Villar *City University London*

David Yeager *Stanford*

Since its inception in 2003, numerous companies have adopted the Net-Promoter Score (NPS) as their main metric for measuring customer satisfaction. The reason for this change is easy to understand. Instead of subjecting one's customers to a range of difficult satisfaction metrics, the NPS promised a much clearer prediction of company revenue growth and customer satisfaction by simply asking customers how likely they were to recommend the company to their friends or colleagues. Then, by simply recalculating the 'likelihood to recommend' responses to the NPS, the companies were argued to be able to predict their future revenue growth. However, the abilities of the NPS have not gone without critique. In fact, the validity of the NPS as a predictor of growth has seldom been investigated in peer-reviewed scientific journals. To rectify this shortcoming, five survey embedded experiments are analyzed in order to evaluate predictive validity of the NPS and the 'likelihood to recommend' question on companies future total revenue growth. Adding to this, the validity of the NPS will be discussed through a survey methodology perspective. Guided by these theories, new versions of the NPS and the survey question are constructed and empirically evaluated. In addition, concurrent validity will be evaluated, comparing both the NPS and the 'likelihood to recommend' question to customer self-reported number of past recommendations. Preliminary analyses of the five survey embedded experiments indicate that there is a linkage between the NPS, the 'likelihood to recommend' question, and future company revenue growth.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

What Participation, Modes, and Costs Indicate about the Feasibility of a New Establishment Survey: The Community Health Policy Implementation Assessment

Douglas B. Currivan *RTI International, Survey Research Division*

Carol Pierannunzi *Centers for Disease Control and Prevention*

Selma Moore *Centers for Disease Control and Prevention*

Chris Stringer *RTI International*

Patty LeBaron *RTI International*

Kevin Wang *RTI International*

Steve Gomori *RTI International*

Devon Wachtmeister *RTI International*

The Community Health Policy Implementation Assessment (CHPIA) will provide the Centers for Disease Control and Prevention (CDC) with tools for a comprehensive surveillance system of local public health policies and practices. Assessing the feasibility of successfully conducting the CHPIA involved collecting data from 120 county health departments in two states (North Carolina and Washington) about their involvement in policies and practices that promote chronic disease prevention and health at the community level. The goal of the pilot was to provide some initial information on the feasibility of collecting these data from local public health officials via two sequential self-administered modes, Web and mail. This poster presentation will use survey data and paradata to examine local public health officials' overall patterns of participation, participation by mode, and the impact on survey costs. Paradata will be used to note participation outcomes at each notification and reminder stage of the survey, to assess the effectiveness of the survey protocol and messages. Comparing overall participation across Web and mail modes will provide evidence on whether both modes are required to meet response rate goals, or whether one of these modes is sufficient to gain participation among local public health officials. The speed and mode of response will also be examined to determine the likely costs needed to launch a national survey of local public health officials using one or both survey modes. The authors will use survey literature on public sector establishments as well as that from general population surveys on sequential Web and mail surveys, to compare with the participation patterns of public health officials in CHPIA. This presentation will contribute to the knowledge base about response patterns in sequential mode surveys among public sector establishments, and especially public health officials.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Bully by Definition: Results of the 2015 National Crime Victimization School Crime Supplement Split-Ballot Experiment

Melissa Cidade *ICF International*

We know that bullying in America's schools is an issue. Examining different sources of data leads to different understandings of the magnitude of the issue. What these estimates obfuscate is that each of these surveys uses a different definition of bullying. To ameliorate the issue of apples-to-oranges comparisons in the rate of bullying victimization, the CDC published a uniform definition of bullying in 2014. In response to the publication of this uniform definition, researchers at the Department of Education changed the way they asked about bullying in school on the 2015 collection of the School Crime Supplement to the National Crime Victimization Survey (NCVS). Since the SCS has been conducted biennially since 1999, researchers had a quagmire: how do they continue to collect the trend data on national bullying estimates while bringing the questionnaire into alignment with the newly published CDC definition? The answer was to embed a split-ballot experiment on the 2015 administration of the SCS. My project is to use the resulting data answer the following research questions: 1. What are the characteristics of students whose reports of bullying are not captured using the CDC Uniform Definition? 2. How much does the national estimate on bullying victimization change when the new definition is in place? I will conclude the presentations with recommendations on how to systematically collect the needed data. I will examine several different considerations in the question construction. The paradigms I am investigating for these recommendations includes a cost/benefit analysis on constructing the question; a duty to the least/hidden curriculum lens on asking about bullying; a consequentialism method of considering how to ask the question in a way that produces the most good and least harm; and a socio-ecological method of asking the question in a way that best embodies the nested nature of the construct.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

How Survey and Big Data Can Work Together? Predicting Customer Engagement Score of Banks and Branches Based on Customer Panel Survey, Company Database and Data from Government Agencies
Dan Yu *Gallup*

This prediction method combines data from multiple sources (i.e. survey, database, government information) to simulate a branch track database of all branches from top 43 banks in U.S. The simulated database can provide the most relevant comparison for a given bank at branch level. A large-scale, probability based web survey was conducted to measure the Customer Engagement Score (CES) of top 43 banks, among which, 10 banks' CESs are known from our database. From these 10 banks, we identify the relationship between CES from survey and database. The relationship is then used to estimate the company level CES for the 33 unknown banks. As some banks' CESs are measured with outbound phone mode, Z-score method is used to first convert all scores from outbound phone to equivalents under web mode. For all 10 banks in our database, we know their branch level CESs, which enabled us to calculate the range of performance (standard deviation) among branches. For an unknown bank, K-Nearest Neighbor (KNN) algorithm is used to estimate its branch level standard deviation based on the known standard deviations from the database. The KNN are decided based on the combination of overall assets, overall deposits, and number of branches. With means and standard deviations for all 43 banks, Branch CES distribution is simulated for all companies, one at a time. In addition, we simulate the exact number of data points that equal to the actual number of branches according to FDIC and FCUA, to reflect the influence of banks based on their branch size in reality. We tested the accuracy of prediction and overall the error is less than 15%. Finally, we discussed limitations and potential directions to improve the prediction accuracy by incorporating other big data sources such as tweets on twitter and Google search results.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

A Demonstration of New Approaches to SEM Model Evaluation with Survey Data

Ana Lucia Cordova-Cazar *University of Nebraska-Lincoln*

Larry Williams *University of Nebraska-Lincoln*

Lok Wa Yuen *University of Nebraska-Lincoln*

Sarah Deng *University of Nebraska-Lincoln*

Ana Lucia Cordova-Cazar *University of Nebraska-Lincoln*

Mengyang Wang *University of Nebraska-Lincoln*

Structural equation modeling (SEM) has become one of the most frequently used data-analysis techniques in psychological research using surveys. Of the various types of SEM models, one of the most popular involves latent variables assessed with multiple indicators, where theories proposing relations among latent variables are tested. Within this framework, it is accepted that evidence of the quality of a model be established before parameter estimates and their significance are attended to. For judgments of the quality of a model, goodness-of-fit measures (e.g., CFI, RMSEA, SRMR) have come to play a dominant role. Of the various goodness-of-fit measures, those that attempt to focus on the viability of proposed relations among latent variables, as compared to those that also incorporate information about relations linking latent variables to their indicators, have been proposed (RMSEA-P, NSCI-P). In this paper we will review recent literature that demonstrates that traditional global fit indices like the CFI, RMSEA, and SRMR fail at identifying model misspecification with latent variable relations. We will also review newer alternative approaches based on the RMSEA-P, NSCI-P, and chi-square difference tests. The use of these newer approaches will be demonstrated using survey data from organizational leadership research. Results will be presented based on data from an organizational survey and a full mediation model proposing that three dimensions of employee justice perceptions influence the quality of their relationship with their leader, which impacts performance of organizational citizenship behaviors. These results will demonstrate how traditional global fit indices focus on the measurement model and are not sensitive to path variable relations and misspecification. The merits of use of the newer RMSEA-P and NSCI-P indices, especially when combined with chi-square difference tests and model comparisons, will be illustrated. The implications of these findings for model evaluation practices as used with survey data will be considered.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Drafting and Wording: Questionnaire Design in Conflict Environments

Hafez Albukari *Yemen Polling Ctr.*

The goal of this paper is to discuss the challenges researchers face when drafting scientifically sound questionnaires in conflict environments. It argues that in such environments great attention must be given to the wording, to ensure that it is precise and at the same time neutral and unprovocative. In Yemen, or many other Arab countries, where conflicts have led to divided societies and religious extremism, certain social norms and habits among the societies make quantitative survey research difficult. For instance, is it difficult to mention certain events and issues part of the conflict in questionnaires. One reason for this is that each party to the conflict has its own vocabulary when speaking of certain events and issues, which support its position in the conflict. This makes it very difficult or nearly impossible to find neutral language. Therefore, one challenge is to avoid a certain bias in the question by using the vocabulary of the conflict parties. More challenges and difficulties arise in questions regarding religious extremism and violence. Like when you ask about prayer places and habits or when you ask about the religious sects or about attitudes toward extremist groups and what are the names of these groups that you are going to use in the questionnaire. Unless researcher be careful, such questions may cause great damages that may end up with unsuccessful/incomplete study; similar consequences are also possible with the tribal traditions contexts. Challenges and restrictions imposed by local conflict environments will be systematically discussed in this paper, while drawing on field experience of the Yemen Polling Center. At the same time, this paper will propose a effective techniques to overcome these challenges.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Who is left off household survey rosters, but picked up by probing?

Sandra Luckett Clark *US Census Bureau*

Household surveys typically begin with the creation of a household roster, which is a list of household members. Creating an accurate list of all household members is a straightforward task in most living situations. However, for some unique living arrangements this becomes more difficult. Inaccurately identifying household members can result in coverage error either by omitting applicable members (under-coverage) or by including inapplicable members (over-coverage). To minimize coverage error, household surveys often include additional probing questions designed to target unique situations. The American Community Survey (ACS) uses four probing questions in the survey's automated modes, which are designed to improve under-coverage by reminding respondents to include household members they may have initially forgotten, and improve over-coverage by removing people who should not be counted as living or staying at the address according to the ACS residence rules. This research uses 2015 ACS data to examine the impact on population estimates of the additional probing questions. Next, it explores the characteristics of the individuals added as a result of the extra questions and the households that had people deleted. Finally, it uses survey paradata from the Internet mode to learn more about the questions, such as how long they took to complete, and whether Internet respondents clicked on a help link while completing these questions. Results from this research indicate that the probing questions used in the ACS impact household population counts and the impact is greater for specific populations. While the questions may improve coverage error, web paradata suggest that they increase respondent burden.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Who You Gonna Call? – Impact of Prenotification Letter Personalization in Establishment Surveys

Brian Orleans *ICF International*
Thomas Brassell *ICF International*
James Dayton *ICF International*
Robert Tortora *ICF International*
Andrew Blevins *The Pew Charitable Trusts*
Theron Guzoto *The Pew Charitable Trusts*
Alison Shelton *The Pew Charitable Trusts*
John Scott *The Pew Charitable Trusts*
Sarah A. Spell *The Pew Charitable Trusts*

Historically, research has supported the personalization of mail communications to respondents in household surveys as a means of increasing response rates (Dillman, Smyth, and Christian, 2014). However, comparatively little research has evaluated the impact of personalizing letters for establishment surveys. With recent gains in sample frame development providing points of contact (POC) at businesses, the impact of personalizing a letter should be evaluated. Personalization of letters for establishment surveys may help bypass gatekeepers; however, it may also lose its effectiveness if the listed respondent is not the targeted respondent. The current study looks to evaluate the effectiveness of three different types of personalization in a RDD CATI survey of small businesses. We addressed letters to one of the following: (1) listed POC, (2) the targeted respondent position (Owner/Human Resource Manager), or (3) a combination (e.g., John Smith or Human Resource Manager). We will compare letter recall (i.e., did the targeted respondent recall receiving the letter), cooperation rates, refusal rates, and overall response rates across the three groups to assess whether any significant differences exist. We hypothesize that condition 1 will yield the highest letter recall, but that no significant differences will be found across dialing performance metrics. Moving forward, this information will be valuable in evaluating the effectiveness of personalizing prenotification letters for establishment surveys, as well as build on the current standing literature regarding its efficacy.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Do Intensified Recruitment Efforts Make a Difference? A Comparison of Early and Reluctant Respondents in a National Surveillance System

Jieru Chen *Centers for Disease Control and Prevention*

Nimesh Patel *Centers for Disease Control and Prevention*

Marcie-jo Kresnow *Centers for Disease Control and Prevention*

With response rates in random-digit-dial surveys declining over the past decades, survey researchers and sponsoring agencies have been intensifying their efforts (e.g., subsampling initial non-respondents for follow-up and offering them greater financial incentives) to recruit participants. These initial non-respondent follow-up strategies are intended to “convert” some initial refusers into respondents, reduce bias, and increase response rates. A higher response rate is generally assumed to be associated with better quality of data. Achieving a higher response rate alone, however, may not necessarily result in higher quality data if data quality is not positively associated with respondent recruitment efforts. Item- and construct-level measurements should also be examined to assess whether response quality is associated with the respondents’ reluctance. Defining early respondents as participants who completed the interview at initial contacts and reluctant respondents as initial non-respondents who were subsequently recruited after additional calls were made and greater incentives offered, this study explores the relationship between reluctance and data quality as well as the effects of reluctance on substantive estimates using a national cross-sectional U.S. adult population survey conducted from 2010 to 2012. Data were analyzed to examine whether differences in individual characteristics between early and reluctant participants help explain any observed variability in population estimates. At the item level, data quality was assessed by examining response patterns. At the construct level, hypotheses were tested separately on a range of key survey outcomes. Preliminary findings show considerable differences in item-level response patterns. While converting some initial refusers into respondents helps increase response rates, response status (early vs. reluctant) is not found to be a significant indicator for a range of population parameter estimates. The implications of findings from this investigation for future data collection efforts and costs are discussed.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

A Sudden Shift in Sampling Strategy: Challenges in Implementation and Communication

Colleen K Porter *University of Florida*

Joshua RTippary *Univ of Florida*

The Florida Consumer Sentiment Index (CSI) telephone survey has been conducted with 500 respondents across the state of Florida every month since 1983. The sample was 100% landline Random Digit Dialing (RDD) until December 2014 and switched to 100% cellphone RDD in 2015. The rate of completed interviews per hour ticked up slightly from .80 to .87, even though time-consuming manual dialing procedures inspired by TCPA were implemented. In some ways, the new sample looks more like Florida: The average age of respondents dropped from 61 to 46 years, more in line with the median age of 41.7 years. The percent Hispanic in our sample more than doubled, from 9.4% to 22.9%, closer to the official 2015 estimate of 24.4% Hispanic. However, the new cellphone sample yielded many more men. In 2014, an average 50.9% of respondents were male, with that ratio switching between genders from month to month within a narrow band. By 2015, the majority were consistently male and by a larger margin, averaging 56.5% in 2015. We have theorized possible reasons and considered various proposals for addressing the gender disparity. For example, we have altered procedures to leave an answering machine message that may encourage female participation if fear of stalking was an issue. In the spirit of the AAPOR Transparency Initiative and because this study's findings are widely reported by media and used by economic analysts, it was important that these changes, which caused a methodological break in the time series, be communicated clearly to those who follow the index. This was done through changing the colors on graphs to clearly indicate a difference in the old and new methods, adding a tab with details on the webpage, and adding explanatory text to the first press release after the change.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

BeHeardPhilly: 1 Year, 8,000 Members, and 17 Surveys Later

Nina Hoe *Institute for Survey Research - Temple University*

Cody Spence *Temple University*

Background: In the fall of 2015, the Institute for Survey Research (ISR) at Temple University launched BeHeardPhilly, the first national citywide survey panel, in Philadelphia. BeHeardPhilly aims to: (1) create a cost-effective and convenient resource for investigators conducting public opinion research in Philadelphia; (2) create a civic-engagement tool for Philadelphians to voice opinions through participating in surveys via their preferred contact mode (web, phone or SMS text), (3) conduct survey experiments and better understand opt-in panel respondent behavior in a local context. Methods: In the last year, BeHeardPhilly has recruited over 8,000 residents, representing every zipcode and subgroup in the city, and has administered 17 surveys on behalf of six City departments and several non-profits. Topics have included rating of city services, knowledge of traffic laws, and opinions of Philadelphia's hosting of the 2016 Democratic National Convention. Results: Findings reveal residents are eager to participate and know survey results, indicating trust in BeHeardPhilly and engagement in city life. Enrollment and participation vary across subpopulations, though. Residents with higher levels of education are both more likely to enroll in BeHeardPhilly and to complete surveys, as compared to those with lower levels of education. Simultaneously, enrollment data indicate that members' preferred contact mode is related to level of education, income, and race/ethnicity. For example, 26% of residents with less than a high school diploma opt to participate via SMS text, compared to 16% overall. Moreover, response rates to specific survey topics are related to other demographics (e.g., males were more likely than females to participate in political surveys, while females were more likely to participate in health-related surveys). Contributions: Findings from the first year of BeHeardPhilly provide insight into survey participation behaviors across subgroups of Philadelphians, and results suggest the need for demographic specific incentivizing to ensure widespread participation and inclusion.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Do incentives during tracking efforts make a difference?

Natalia Ibanez *Decision Information Resources*
James Cooper *Decision Information Resources Inc*
Leslyn Hall *Redstone Research, LLC*
Jo Anna Hunter *MDRC*

Achieving response rate goals in longitudinal studies are more likely when an emphasis is placed on vigorously tracking participants (Cotter, et al., 2005). The specific tracking strategies are contingent upon studies' characteristics, such as length of follow-up, cost, and population (Hunt and White, 1998). Using data from an antipoverty evaluation, the proposed study will examine a multi-mode tracking strategy implemented prior to a 30-month post random assignment follow-up. Specifically, 4,750 sample members were sent a mailed flyer at 6-months post random assignment (and a survey-branded magnet), again at 12-months, and then every quarter until the survey period began. At each quarter, members were asked to provide updates and confirm their contact information. To determine how a small incentive would impact responsiveness, sample members were divided into 4 different groups – three treatment groups and one control – each receiving a different incentive at different quarters. Two treatment groups received a \$2 pre-incentive, one group received two \$2 pre-incentives and the control group received none. The proposed study will explore the effects of the tracking variations on the number of updates and confirmations received. Further, researchers will examine whether tracking groups revealed an influence of pre-incentives on survey responsiveness and if updating/confirming contact information – regardless of tracking group – influenced response rates. Finally, researchers will explore whether the tracking incentives reduced data collection costs (measured through number of days to complete and number of touches).

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Modeling Follow-Up Survey Completion in an Adolescent Cohort Using Baseline Survey Predictors

Robert Tortora *ICF International*

Joanne Delk *University of Texas*

Melissa B Harrell *University of Texas*

Cheryl L Perry *University of Texas*

The Texas Adolescent Tobacco and Marketing Surveillance System (TATAMS) began following a longitudinal cohort of 3,907 6th, 8th and 10th grade students in the 2014-2015 school year. Participants were recruited from and took the baseline survey, varying in length from 117 to 337 questions, at school. Those who completed the wave 1 survey were given a \$10 gift card upon completion. Follow-up surveys occur roughly every 6 months and participants receive links via email, text, and conventional mail to complete the survey on their own time. Response rates were 63.7% and 70.0% for waves 2 and 3, respectively, with 75.2% completing at least one follow-up survey. Logistic regression models were created to determine which student-level baseline variables were predictive of not completing the wave 2 and 3 survey. The strongest predictor of not completing any follow-up surveys was not providing an email address at baseline (AOR 7.0, 95% CI: 4.6, 10.8). Additionally, those that did not complete follow-up were more likely to be male, in 10th grade compared to 8th, White/other versus Hispanic, of a higher socio-economic level, failed to redeem their wave 1 gift card, be a past 30 day tobacco user, provided zero or one phone numbers at baseline, and had greater than 180 days between baseline and invitation to wave 2. The number of questions received on the baseline survey, number of item nonresponse, being susceptible to tobacco use and being an ever tobacco user were not predictive. This model can help inform TATAMS and other cohort studies with youth on how to better allocate study resources, target those at risk of not participating in the future, and to maintain or improve response rates.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Can I Get Your Attention Please? Effects of Using a Question in the Email Subject Line to Improve Survey Response Rate

Erin Czyzewicz *SSRS*

Robyn Rapoport *SSRS*

Eva Chiang *George W. Bush Institute*

Catherine Jaynes *George W. Bush Institute*

As more research is being conducted online and respondents are increasingly recruited via email invitations, survey and market researchers have begun to investigate ways of making email subject lines more inviting since they are key to engaging respondents and inducing them to open the email rather than deleting it. The poster reports on an experiment done in 2016 as part of a survey of public school principals on behalf of the George W. Bush Institute's Education Reform Initiative. The objective of the survey was to gain a better understanding of principals' perceptions regarding their roles in ensuring students' academic success as well as related issues including how principals perceive their interactions with parents and others in the school community. The experiment tested two e-mail subject lines for invitations: • Running Your School: Share Your Expertise • Running Your School: What Challenges Do You Face? Our hypothesis was that the subject line with the question text would be more intriguing for principals and would be associated with a higher response rate. Notably, some market researchers suggest that asking a compelling question in the subject line can engage people in the issue. We tracked all resolved statuses, that is, completed interviews along with those who opted out, didn't qualify and/or didn't consent to participate in the survey. We also tracked other paradata, such as whether the email was opened and if the survey link was clicked on and compared all of these variables across the two experimental conditions. Our analysis will show that, overall, the invitation question text in the context of our survey was associated with a somewhat smaller number of opened emails and completed interviews, and a significantly smaller number of resolved cases. Interestingly, the biggest differential is for principals living in the south, possibly pointing to a region-specific response.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Changing Immigration laws and their Impact on Attitudes within Qatar's Sponsorship System

Abdoulaye Diop *Qatar University Social and Economic Survey Research Institute (SESRI)*

Kien TLe *Qatar University Social and Economic Survey Research Institute (SESRI)*

John Lee P Holmes *Qatar University Social and Economic Survey Research Institute (SESRI)*

Semsia Al-Ali Mustafa *Qatar University Social and Economic Survey Research Institute (SESRI)*

Buthaina ALKhelaifi *Qatar University Social and Economic Survey Research Institute (SESRI)*

Noof Al-Rakeb *Qatar University Social and Economic Survey Research Institute (SESRI)*

This paper analyzes the variation in ways that different communities in Qatar perceived the likelihood and effects of changes to the state's immigration laws. Criticism of GCC countries' labor laws from international regimes has become focused on Qatar in the context of its hosting of the upcoming 2022 FIFA World Cup. This has led to external pressure to change or eliminate the "Sponsorship" legal framework that regulates all non-citizens working in Qatar. Qatar University's Social and Economic Survey Research Institute (SESRI) has been measuring changes in expectations regarding workers' rights and the sponsorship system via phone and in person interviews since 2010. After an announcement last year that changes to these laws were forthcoming and with the promulgation of the labour laws on December 2016, the survey followed up with a battery of questions measuring understanding of and beliefs about whether changes will be meaningful once implemented. We measure the degree of support by nationals and other groups for the new changes, including how these changes will affect their businesses, social fabric and expectations for an improved international reputation. In addition to the social capital projects, the quarterly surveys also encompass a series of questions about consumer expectations and multiple quality of life measures. The surveys are conducted in seven languages covering the country's three major socio-economic groups: Qataris, high income expatriate ("white collar") and low income migrant ("blue collar") workers. As a consequence we are able to use multivariate analysis to understand underlying relationships between attitudes and support for or opposition to the system and its modification. This study illustrates the power of survey research to help better understand complex social processes generally, and labor migration in particular, issues of increasing importance in the current world economy.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Multimodal Evaluation of Gambling Attitudes in Population of Louisiana

Steven J Dick *Picard Center for Child Development and Lifelong Learning*

Raymond WBiggar *Picard Center for Child Development and Lifelong Learning*

Irv Esters *University of Louisiana at Lafayette*

The effect of gambling on a population can range dramatically from a vast influx of money to the local economy to personal financial losses or even bankruptcies. Consistent with national trends, Louisiana is treating problem gambling as a public health issue. Currently, Louisiana's 4.7 million residents are served by four horse racing tracks, 15 riverboat casinos, three Indian casinos, one major land-based casino and nearly half of the state's parishes allow gambling establishments to operate. As of June 2014, 1,945 video poker outlets offer gaming throughout the state. An important step is to understand the scale of the problem. The Louisiana Department of Health has commissioned two surveys that can provide insight into the attitudes and potential risks of gambling in the state. First, two regionally stratified samples of adults (2008 and 2016, N=2,700 each) investigated gambling attitudes with a measure of possible problem gambling behaviors. Second, the Caring Communities Youth Survey (CCYS) sought a biennial census of 6th, 8th, 10th, and 12th-grade students designed to look at a broad spectrum of risk and protective factors for the state youths from 2002 to 2016 (N=92,000 in 2014). This study seeks to triangulate adult and youth perceptions of gambling experiences, behaviors, and attitudes from the 2008 and 2016 surveys. Researchers will be able to investigate differential effects based on demographics, and regions (e.g., urban/rural, availability of gaming establishments). In addition, the 2008 survey will include a substantial portion of the population still displaced by hurricanes Katrina and Rita. The result should reveal valuable information about gambling in this mostly rural state and the interaction of youth and adult populations.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Non-Contact and Data Quality in the American Time Use Survey

Dawn V. Nelson *U.S. Census Bureau*

Beth Ashbaugh Capps *U.S. Census Bureau*

The American Time Use Survey (ATUS) is sponsored by the Bureau of Labor Statistics and administered by U.S. Census Bureau telephone interviewers. At the survey's inception in 2003, the response rate was just below 60%; and now, thirteen years later, the response rate is below 50%. The survey's refusal rate has hovered around 20% each year; however, the noncontact rate has shown a marked increase, from single digits in early 2000's to nearing 30% in recent years. Should survey practitioners put forth additional efforts to convert refusals and track down non-phone cases? Will these additional efforts matter – will data quality be improved? Previous researchers have demonstrated that ATUS data quality (measured by the total number of reported diary activities) actually worsens when the nonpropensity score increases (Fricker and Tourangeau, 2010) and suggest that survey organizations may want to divert resources away from recruitment of difficult respondents and focus on other error-reduction techniques. What about those cases we did not contact because we could not locate a current telephone number? Despite multiple attempts by mail, the case remained a non-contact at closeout. Do these nonrespondents resemble the low-propensity responders? What additional efforts should we take to complete an interview? Using CPS data and ATUS paradata, we examine ATUS noncontact cases. We explore data quality with CPS measures and look at typical ATUS non-contact households and sample persons. We conclude with recommendations for survey practitioners.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Response Heaping in Weight and Height Self-Reports: Consequences for Predictive Validity

Marina Stavrakantonaki *Univ of Illinois at Chicago*

Allyson Holbrook *University of Illinois - Chicago*

Response heaping (also sometimes called rounding or digit preference) occurs when respondents answering a question that requires a specific numerical response (e.g., How many times in the past 12 months have you gone to the doctor?) give answers that are divisible by 5 (e.g., 5, 10, 15). In this research, we examine response heaping in respondents self-reports of their weight and height using data from the National Health and Nutrition Examination Survey (NHANES). Specifically, we first examine the prevalence of heaping and predictors of heaping in weight and height reports. What are the characteristics of respondents who give heaped responses? For example, past research has found that respondents who weigh more are more likely to give heaped values. Second, we examine whether heaping is associated with inaccuracy in weight reports by comparing self-reported weight to weight measured objectively as part of the NHANES and the predictors of the accuracy of weight self-reports. Previous evidence suggests, for example, that respondents who give heaped responses consistently underreport their weight and that women are more likely to do so than are men. Finally, we examine whether potential inaccuracy introduced by heaping affects the predictive validity of self-reported weight. In order to do so, we compare the predictive validity of respondents' BMI using self-reported and measured weight and height (for respondents who gave heaped self-reports and those who did not) in order to test whether the self-reports of respondents who gave responses that were divisible by 5 showed lower predictive validity than the self-reports of respondents who gave responses that were not divisible by 5. Specifically, we examine the extent to which BMI predicted physical conditions like high cholesterol, hypertension, type-2 diabetes, cancer, breathing disorders such as asthma and sleep apnea, gallbladder disease, heart disease as well as mental conditions like depression and anxiety.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

A Big Data Perspective of the 2016 Presidential Election

Brent SH Waddington *Resonate*

Dan Scantlebury *Resonate*

Kevin Shea *Resonate*

Resonate uses proprietary survey data and petabyte-scale internet behavioral data (website visits and URL topics derived from natural language processing techniques, totaling billions of records) to develop supervised machine learning algorithms that can impute survey responses across approximately 200,000,000 internet devices located within the United States. Because the algorithms estimate the probability of how any internet user would have answered a survey question and we can also assess the location of their device with some level of granularity, our methods provide a unique Big Data perspective of the 2016 Presidential election. With this in mind, we ran a survey between August 4 and September 1 of 18,480 respondents who were asked, among other topics, whether they intended to vote as well as who they would vote for (out of major party candidates and a selection of Third Party contenders). Using panelists responses as labels and their browsing behavior as features, we are able to develop accurate predictive models using standard techniques - we randomly partitioned 20% of the survey and behavioral data into a final validation set, and used a cross-validation procedure on the remaining 80% - the training set - to tune model hyperparameters so as to avoid overfitting (e.g., "memorizing" training data), and assessed model accuracy on the validation set. Using this methodology and a simplistic "algorithmic vote" within each state, we correctly predicted the Electoral College in 48 elections - missing only Minnesota, Pennsylvania, and Florida, and all by razor thin margins. Contrary to punditry, our results confirm the power of Big Data to empower campaign decisions.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Modeling Support for Tolling in the Lone Star State

Chris Simek *Texas A&M Transportation Institute*

The traditional means of funding transportation infrastructure in the United States (U.S.) is the Federal Highway Trust Fund, which relies heavily on the Federal Fuel Tax for revenue. Due to a myriad of reasons, including, but not limited to, increased passenger vehicle fuel efficiency, the rise in popularity of alternative fuel vehicles, and not being linked to inflation, the Federal Fuel Tax can no longer support the levels of funding necessary to meet growing demands placed on the transportation network. From 2010 to 2015, the state of Texas population increased by 9.2 percent, making it the second fastest growing state in the U.S. during this time period. Similarly, in 2015, Texas had the second highest gross state product (GSP). The demands placed on the state highway network by a growing population and a growing economy are significant. As a result, Texas has become increasingly interested in alternative transportation funding mechanisms, including tolling. The 2016 Texas Transportation Poll results suggest that, while nearly three-fourths of Texans believe there is a need to increase statewide transportation funding, there is little support for either building more toll roads (28 percent support) or increasing the tolls on existing toll roads (14 percent support). Regression techniques were used to identify, if attitudinal and behavioral attributes collected in the 2016 Texas Transportation Poll could be used to predict support for tolling. The results suggest that variable such as support for private sector investment in transportation infrastructure and registered voter status are the most significant common predictors. Results from this presentation will add to the growing body of research available to transportation planning professionals regarding the use of controversial and often mis-understood mechanism for funding transportation.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Exploring Philadelphian's Perceptions of Police Presence, Approachability, and Conduct Using an ABS Sample and the BeHeardPhilly opt-in Panel

Heidi Grunwald *Temple University*

Background: On October 20, 2016 Temple's Institute for Survey Research (ISR) was commissioned by the Mayor of Philadelphia to field the Philadelphia Resident Survey (PHL-RS) to better understand Philadelphian's perceptions of the quality of their municipal services such as public safety, streets, parks and recreation, economic development and health and human services. In fall of 2015 ISR created and began recruiting Philadelphia residents into the first citywide opt-in survey panel called BeHeardPhilly (BHP). Philadelphia is the poorest of the large cities in the country and a difficult city to use mail or dual-frame designs due to cost and low response rates. We decided to create an opt-in panel to maximize the value of the local context and to increase civic engagement. As of November 10, 2016, BHP has over 8000 Philadelphia panel members who are eligible to be selected for research studies. Methods: The PHL-RS was fielded in English and Spanish using 1) an address-based sampling frame of 4500, 2) the eligible members from the 8000 member panel and 3) in-bound recruitment of respondents via a press release to media outlets. The ABS sample was mail only. BHP panel members who had elected SMS as their preferred mode were sent an SMS alerting them that the survey was too long for SMS and they were delivered a web link. The in-bound opt-in respondents (a technique used by many small non-profits) were asked to use a web-link. This paper will explore the differences in Philadelphian's perceptions of the police presence, responsiveness, approachability, ability to prevent crime, and police officer conduct. Early results suggest statistically significant differences in estimates by demographic groups, types of samples as well as modes across sample types. Contributions: To better understand respondent behaviors of Philadelphia residents and explore differences in estimates by sample type and mode.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Negativity Bias: The Link between Communication Strategy and Hostility in Congress

Adam Hughes *Pew Research Center*

Solomon Messing *Pew Research Center*

Patrick Van Kessel *Pew Research Center - Washington, DC*

Partisan hostility in the U.S. Congress is negatively associated with bipartisan cooperation, legislative productivity, and public trust in government. By examining hostility and conflict in Congressional communications, we examine the ways that polarized lawmakers communicate with the public. Using supervised text analysis, we examine hostility across Congressional press releases and Facebook posts issued during the 114th Congress, a sample of more than 250,000 individual documents. We find a link between attention-seeking and hostility: party leaders and polarized members seeking to influence policy agendas are more likely to attack and press releases, which require media attention to reach constituents, are more likely to contain hostility than Facebook posts. Our results show that hostility does indeed generate more attention, at least in the context of engagement with Facebook posts, despite the fact that the public reliably states a strong desire for legislative cooperation. We find that posts containing political conflict earn more likes, shares, and comments than those focused on constituent benefits or bipartisan legislative action. Furthermore, attention-seeking members tend to attack other branches of government - most notably, the presidency, but also bureaucratic agencies and the courts. Our findings help understand how individual members' communications strategies can increase discord in Congress and exacerbate inter-branch conflict.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Nonresponse Analysis and Field Protocol Recommendations The AAPOR Annual Membership Survey

Chase Harrison *Harvard University*

Kristen Olson *University of Nebraska - Lincoln*

Anna Wiencrot *NORC at the University of Chicago*

Morgan Earp *Office of Survey Methods Research*

HarmoniJoie Noel *American Institutes for Research*

Caitlin Deal *American Institutes for Research*

Organizational surveys offer special challenges and provide distinct opportunities to analyze nonresponse and tailor field protocols to optimize response rates among different groups. This paper uses the Annual Membership and Conference Attendee Survey of the American Association for Public Opinion Research (AAPOR) as a case study, and discusses nonresponse patterns across different groups and presents ideas for tailored field approaches to increase response rates across different groups. The AAPOR membership survey is conducted annually and targets a census of members in the organization as well as non-members who attend the organization's annual conference. In recent years the survey has targeted approximately 2,000 members and obtained response rates of about 30%. As part of an organizational commitment to employ research best-practices, the AAPOR Membership Committee has analyzed nonresponse across the surveys and is planning targeted field-protocols to attempt to address differential nonresponse. This paper presents our nonresponse analysis for the 2015 and 2016 AAPOR surveys. In general, our analyses did not find significant differences on response rates based on age, education and gender, while white-non-Hispanic members were slightly more likely to respond compared to others. Members who have attended multiple conferences are more likely to respond compared to those who have attended fewer, and the number of years in a member had been employed in the research industry was also positively correlated with response. Importantly, student members, younger members, new members, and members from commercial research organizations were less likely to respond to the survey than others. We discuss pilot efforts begun in 2016 to increase responses among these groups, and present further ideas and plans on how to increase overall response and decrease differential nonresponse to this continuing survey.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Reducing the Number of Pages of a Mail Questionnaire in a Mail/CATI Mixed Mode Survey: How Does It Affect the Response Rates and Costs of Each Mode?

Michelle A Cantave *IMPAQ International*

Eban Lewis *CSS Research*

Amy Djangali *IMPAQ International, Inc.*

Previous research examining the effect of reducing the perceived length of paper questionnaire by reducing the number of pages through formatting, without altering the questionnaire content, is mixed. In addition, to the best of our knowledge, no prior work has examined whether reducing the perceived length of the paper questionnaire in a mixed mode study generates higher response rates and reduces costs for both modes, not just the mail component. The Fee-for-Service Consumer Assessment of Healthcare Providers and Systems Survey (FFS CAHPS) is a sequential, mail, then centralized telephone follow up mixed mode design. From 2011 to 2015, the FFS CAHPS paper questionnaire was a 16 page booklet with additional instructional pages. In 2016, we reduced the booklet to 12 pages and simplified the instructional language (the number of questions remained the same). Confirming the previous literature, we observed an increase in the number of mail surveys received. Though the increase cannot be fully attributed to the reduced pages, no other large-scale changes were made to the survey that would account for such a sharp increase in the mail response and for a halt to the previously downward trend in the response rate. We will present our findings on the increased response rates and decreased costs in both modes, as well as an analysis of the response rate increases by demographic characteristics.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Development of a multi-mode tobacco surveillance system for use in sub-Saharan Africa

Emmanuela Gakidou *University of Washington*

Farah Daoud *University of Washington*

Miranda Bryant *University of Washington*

Stanley Kamande *University of Washington*

Marissa Reitsma *University of Washington*

Aubrey Levine *University of Washington*

Current tobacco surveillance methods in sub-Saharan Africa are time consuming and expensive, leading to limited implementation, inconsistent geographic coverage, and an absence of reliable time-series data. As tobacco use in Africa rising and the market is expanding, the burden of disease is shifting toward non-communicable diseases. An improved tobacco surveillance system is needed. We aimed to develop an approach that is low-cost, timely and reliable, to monitor tobacco use trends. We conducted a global landscape assessment to identify existing data sources, instruments, tobacco and non-tobacco surveillance systems, technologies, and factors specific to the African context. We consulted multi-disciplinary experts in surveillance, survey methods, social media analysis and tobacco use in Africa prior to the implementation of an on-the-ground pilot in Nairobi, Kenya. The resulting surveillance system utilized a multi-mode data collection framework to increase efficiency and timeliness consisting of social media, observation, short message service (SMS), in-person surveys, and biomarkers. Data sources were leveraged in an attempt to account for coverage and response biases. Pilot results were validated through a comparison to nationally representative surveys in Kenya.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

To probe or not to probe, that is our question: The impact of probing in final response distributions for telephone surveys in global contexts

Sofia Pinero Kluch *Gallup*
Stephanie Marken *Gallup*

Survey researchers have long asserted that probing is important in interviewer-administered surveys when respondents report they don't know or refuse to answer a question. Probing has been proven important in increasing the rate of valid responses, and many suggest that final responses more accurately reflect respondents' truest opinions. Still, probing places a burden upon interviewers and can increase the total survey length. Some research suggests it can lead to poor quality data to respondent guesswork when asked knowledge questions (Sanchez, 1992). However, little data exists to demonstrate the impact of probing on other types of questions such as factual data about the respondent or attitudinal questions. For example, are factual questions such as "do you have health insurance coverage, yes or no?" as susceptible to probing as attitudinal items are such as, "do you approve or disapprove of the leadership of the European Union?" Further, little is known about how probing impacts the quality of data in different cultures and countries. To address these questions, Gallup conducted an experiment in the U.S., Germany, and Norway using the Gallup Daily Tracking survey and the Gallup World Poll surveys to quantify the impact of probing. Gallup probes initial don't know and refusal responses once before accepting them as volunteered responses. For a limited time, Gallup asked interviewers to note if they had to probe respondents at specified questions (the U.S.); asked interviewers to record which questions they probed across the full survey (Germany); and asked interviewers to record probing for all survey items so that each item could be examined along with its final coded response (Norway). In this presentation Gallup shares experiment results including how item distribution is impacted by probing, how different types of questions are differentially impacted, and the types of respondents who are most frequently probed.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Small Business' Reactions to Auto-IRA Plan Features

Andrew Blevins *The Pew Charitable Trusts*
Andrew Blevins *The Pew Charitable Trusts*
Theron Guzoto *The Pew Charitable Trusts*
John Scott *The Pew Charitable Trusts*
Alison Shelton *The Pew Charitable Trusts*
Sarah Spell *The Pew Charitable Trusts*
Thomas Brassell *ICF International*
Robert Tortora *ICF International*
James Dayton *ICF International*
Brian Orleans *ICF International - Fairfax, VA*

As policymakers look for ways to increase retirement security, many are considering auto-IRA plans, in which employees without an employer sponsored plan are automatically enrolled into an IRA that is managed by a third-party organization such as a financial firm or government. Though these plans are meant to mitigate employers' roles beyond processing enrollment and contributions, their support is important to the success of these plans. Understanding what drives business support or opposition would enable policymakers to tailor both the policies and the presentations of these policies to effectively increase retirement security. Using data are from a nationally representative telephone survey of small, private-sector businesses, we examine how employers without retirement plans react to auto-IRA plans depending on plan features. We specifically examine how varying contributions rates are related to support and opposition to auto-IRA plans. Additionally, how support changes depending on if the auto-IRA plan is sponsored by the federal government, state government, a mutual fund company, and an insurance company. Based on research on employees' attitudes, we expect that employer support would not differ when presented with a 3 percent or 6 percent default employee contribution. Further, focus group research suggests that a distrust of government would mean employers would more likely support auto-IRA plans when they are sponsored by a private company than a government entity. Preliminary analyses show businesses seeing an auto-IRA proposal with a 6 percent default employee contribution are not statistically significantly different from those seeing a 3 percent contribution. Enacting policy with a higher contribution rate—and, thus, higher savings rates—is likely not to deter businesses. Employers had significantly less support for government sponsored plans, indicating the policies should be presented with clear language that third-party organizations will be those managing funds and only employees could access their money—not the government.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

App-Based Diary Studies as an Alternative to Traditional Usability Testing for Mobile Applications

Ashley Schaad *ICF International*

Arlen Rosenthal *ICF International*

James Dayton *ICF International*

Daniel Baker *MFour Mobile Research*

Alec Schaefer *MFour Mobile Research*

Usability testing for mobile applications is an important component of ensuring their adoption by the target audience, yet it is a step that is commonly skipped – particularly for small scale apps and those with a low incidence target audience. This may be due, in part, to the time and cost restrictions of conducting traditional laboratory-based moderated usability testing. Unmoderated testing options cut down on the researcher's time and the costs associated with laboratory testing, but unmoderated testing generally involves screen and voice/video recording that requires specialized software to capture and results in data that can be time-consuming to analyze. In this paper, we describe two mobile app usability tests conducted in a diary study format using an app-based survey panel with national representation. The tests were unmoderated and consisted of a series of daily testing scenarios with follow-up questions during a one-week testing period. No screen recording software was used; screen shots were used to confirm the respondents had successfully completed the tasks. The follow-up questions contained primarily closed-ended questions that required little time to analyze. For the open-ended questions, respondents were randomly assigned to enter either text-based responses or to record video responses which were then converted to text for analysis. This allowed us to further examine the benefit – or lack thereof – of video-based responses as opposed to standard text-only methods.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Studying Veterans' Post-Service Transition Experiences in an Era of Declining Survey Participation: Designing and Fielding The Veterans Metric Initiative (TVMI)

Bradford Booth *ICF International*

Patricia Vanderwolf *ICF International*

Dawne Vogt *National Center for PTSD*

Daniel Perkins *Pennsylvania State University*

Ensuring that Post-9/11 military veterans experience a successful transition from military service to civilian life has become a major policy goal across the federal government and for the nation. Federal agencies, state and local governments, non-profit organizations, and the commercial sector have been active in developing and implementing programs and services to support this goal. However, there are few rigorous studies that measure the effectiveness of these programs on new veterans' transition over time, and no longitudinal panels exist that track the transition experiences of a large, representative cohort of veterans in ways that can link program participation with outcomes (Mattox and Pollard, 2016). This gap is being filled through a unique public-private research collaboration led by the Center for Public-Private Partnerships at the Henry M. Jackson Foundation for the Advancement for Military Medicine, Inc. (HJF), titled The Veteran Metrics Initiative (TVMI): Linking Program Components to Post-Military Well-Being. The TVMI study is using a VA-provided list frame to recruit 7500 recently transitioned veterans to participate in a longitudinal survey that gathers data on their overall well-being across several dimensions (i.e., vocation, economic, health, social relationships), as well as the extent and level of their participation in various support programs (e.g., community-based, governmental, and online). The TVMI team, which includes investigators from HJF, VA, DoD, Pennsylvania State University and ICF, faced a number of significant methodological challenges when considering the best study design, including decreasing response rates to surveys, the need to ensure representation from all subgroups of separating veterans, and regulatory barriers to data sharing between federal and non-federal collaborators. We provide an overview of the final study design and results of a pilot to test response rates to three different mailing conditions, and we present data on the characteristics of the sample recruited for the baseline wave.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

The Risk of Place: A Spatial Analysis of Tobacco Usage and the Density of Tobacco Retail across Population Subgroups

Jennifer Cantrell *Truth Initiative*
Andrew Anesetti-Rothermel *Truth Initiative*
Ned English *NORC at the University of Chicago*
Peter Herman *NORC at the University of Chicago*
Ilana Ventura *NORC at the University of Chicago*
Morgane Bennett *Truth Initiative*
Michael Halenar *Truth Initiative*
Elizabeth Hair *Truth Initiative*
Donna Vallone *Truth Initiative*

Research has shown that accessibility to retail tobacco (i.e., tobacco retailer density) influences consumption patterns. Most research has calculated tobacco retailer density as per unit of area or population, however, without considering subgroups such as those defined by poverty, education, or race/ethnicity. There thus exists the potential for studying the relationship between tobacco retailer accessibility and the behavior of specific at-risk groups. We use results of a study integrating data from geocoded tobacco retailers; extant demographic and health information; and survey data from the Truth Initiative's TLC Cohort to provide a nuanced analysis of the impact of retail density on smoking behavior and attitudes. An adaptive-bandwidth kernel density estimation (KDE) approach assessed how the density of tobacco retailers varies across population subgroups. Then, to understand differences in the accessibility of retail tobacco between minority and non-minority smokers, we utilized a static-bandwidth KDE approach using the TLC Cohort, a nationally representative sample of youth aged 15-21 years old. Overall, our results show that analyses based on definitions of space can vary depending on the analytical approach and the units themselves, and so it is necessary to have a theoretical and methodological basis for analysis. Furthermore, we illustrate how GIS and geospatial methods can help integrate survey and extant data in models. This study is valuable to researchers interested in incorporating geospatial data into the analysis of survey data.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Implementing Optical Mark Recognition Technology to Process Mailed Surveys at a State Cancer Registry

Natalia L Herman *New Jersey State Cancer Registry - Cancer Surveillance Research Program*

Steve Joslin *Gravic, Inc.*

Carolina Lozada *Rutgers Cancer Institute of New Jersey*

Lisa E Paddock *New Jersey State Cancer Registry*

Jennifer Tsui *Rutgers Cancer Institute of New Jersey, Division of Population Sciences*

Antoinette M Stroup *New Jersey State Cancer Registry*

The New Jersey State Cancer Registry (NJSCR) is a population-based registry that collects data on all cancer cases diagnosed and treated in New Jersey. Additionally, this central cancer registry conducts epidemiologic surveys of cancer amongst the state's residents. The NJSCR-CSR conducts an average of 10-12 survey research studies annually, contacting approximately 3600 cancer patients in the process. Studies frequently consist of mailing paper surveys (22-40 pages), both one-time surveys and longitudinal studies with multiple surveys per participant. To improve and expedite data entry and to maintain quality control (QC) the NJSCR-CSR piloted the use of an optical mark recognition tool (Remark Office OMR) in a large survey research study conducted in collaboration with the Rutgers Cancer Institute of New Jersey (RCINJ), "Improving Patient Access to Quality Cancer Treatment" (IMPACT). Breast, prostate, colorectal, cervical cancer survivors were recruited to complete a one-time site-specific survey by mail. Templates were created in Remark Office OMR for each of the site-specific surveys. CSR staff implemented barcodes on returned, completed surveys, scanned them using a desktop scanner, and saved the data on a secure network server. Data were read into the software using pre-scanned surveys; staff manually entered handwritten responses and corrected any exceptions. CSR staff performed quality control of 274 surveys. Over 700 surveys were scanned for primary data entry within 5 months, but over 60% were completed within a 2-month period with 3 staff working up to 95% on the project. Surveys had up to 500 variables to be reviewed, including handwritten responses. While NJSCR-CSR does not usually perform data entry for collected surveys, we will be offering it as a service for investigators using state cancer registry data. We will also be looking into expanding its uses beyond surveys for additional study documentation.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Extracting Extant Information from Digital Photographs: Tobacco Retail Establishments

Robert Montgomery *NORC, Univ. of Chicago*

Andrew Anesetti-Rothermel *The Schroeder Institute at Truth Initiative*

Jennifer Cantrell *Evaluation Science and Research at Truth Initiative*

Ned English *NORC at the University of Chicago*

Peter Herman *NORC at the University of Chicago*

Yongheng Lin *NORC at the University of Chicago*

Advances in data acquisition, transformation, and analysis tools are opening up the possibility of incorporating previously untapped data sources into social science research. In the area of tobacco control research, the availability of tobacco products has been shown to affect tobacco usage. The presence of products aimed at younger demographics such as menthol cigarettes, little cigars, and cigarillos are of particular concern. However, retail outlet level product availability data is difficult to come by; to create such data, we applied machine learning via the TensorFlow platform on a library of digital photographs of store displays containing tobacco products in West Virginia, USA. A subset of the image library was classified manually by product displayed to act as a training set for a classification model, while a second subset was employed as a validation set to evaluate model performance. The remaining photographs were classified programmatically with this model, yielding a dataset indicating the presence of different tobacco products by store. In this poster, we will describe the data extracted from the image library, how it compares to a manual review, and how it related to expectations from other sources, including Nielsen scan data. Our research illustrates the challenges associated with image classification but also the utility of extract data for social science research.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Utility of Party Identification (ID) in Political Polling

Robert Benford *GfK*

John Lien *GfK*

Ge Tang *GfK*

Frances Barlas *GfK*

Polling in the 2016 Presidential election proved problematic for the polling industry. From polls to poll-aggregators, probabilities indicated a near certainty for the Democratic candidate, Hillary Clinton. While the post-mortem will most likely identify many reasons for this, and polling will yet go through more evolution, the AP-GfK Poll can shed some light on possible noise in voting behaviors. The poll in the weeks before the 2016 election showed a substantial lead for Hillary Clinton. While she did win the popular vote, it was closer than polls estimated. Even though there was outrage at Donald Trump's fitness for the Presidency he won the election. Were there latent clues that this outrage was a politically correct front affecting polling estimates? Using a rich battery of profile data for members of KnowledgePanel – nearly 2,000 variables secured from a dozen different surveys – and the AP-GfK Poll statistics, we will show that Party ID was in flux among panel members close to the election as compared to what was reported months before.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Evaluating non-response to cognitive functioning measurements in a federal survey

Debra J Brody *National Center for Health Statistics, CDC*

Ellen Kramarow *National Center for Health Statistics, CDC*

It is widely recognized that advanced age and poor health status are associated with lower survey response rates. Using data from a multi-purpose federal survey that included an interview and health examination, we evaluated response rates for the cognitive functioning component by age and other socio-demographic and health characteristics. During the 2011-2014 National Health and Nutrition Examination Survey (NHANES) cognitive performance was assessed among adults, 60 years of age and older, using three standardized instruments. In addition, questions on self-perceived memory and cognition were asked during the home interview. The overall response rates for completion of all three cognitive performance assessments was 89.5%. In addition to age (older) and self-reported health status (poorer), non-respondents also differed by survey cycle, race and Hispanic, birth place, education and other characteristics, including self-perceived cognitive function. Assessment scores were evaluated before and after re-weighting and nonresponse adjustment. Discerning characteristics of non-respondents is useful for a complete assessment of cognitive performance among older adults in the United States household population, and to determine necessary adjustments or reweighting to minimize possible biased estimates for this important public health concern.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Propensity Score - Post Stratification or Sample Frame

Kalina Popova *Nielsen*
Alyson Lamberti *Nielsen*

This project evaluates the effect of moving the propensity score from the post stratification stage to the sampling frame. Nielsen's Consumer Insights division was among the pioneers of using propensity scoring to reduce the bias inherent in online survey samples. Traditionally we calculate the propensity score on the back end, after data collection, and do so on each online sample individually. An alternative approach would assign a periodically updated propensity score to the entire panel and then use this as a variable in the sampling frame of each study going into field. The possible benefits of this alternative approach are operational (time savings as the propensity scoring is done once for the entire panel instead of on a survey by survey basis) and methodological (more representative samples as attitudinal and behavioral skews captured by the propensity score are addressed up front, as well as in the data weighting to correct for response rate fluctuations). We will evaluate the alternatives by running two simultaneous surveys, one sampling as usual and assigning a propensity score on the back end and another where we including the panel score in the sampling frame. We will evaluate the resulting representativeness of the two samples, the weighting efficiency and closeness of each sample responses to known external benchmarks.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Increasing the Efficiency of Computer Assisted Telephone Interviewing

Jordan Misra *United States Census Bureau*
Rachel Horwitz *United States Census Bureau*
Aliza Kwiat *United States Census Bureau*
Beth Newman *United States Census Bureau*

Although telephone interviewing has become less reliable and more costly in recent years due to increased mobile phone usage, caller-identification, and other technologies, it remains one of the primary modes of data collection for many surveys. Since these technological advancements can have varying impacts, there is a need to examine how calls can be made more effectively. This research focuses on increasing the efficiency of CATI in the context of a mixed-mode survey. The goal of this study is to identify strategies that focus CATI efforts on respondents that are most likely to complete the interview in that mode. Using a discrete time to event model, we analyzed historical data from the Current Population Survey to determine how contact attempt characteristics influence the likelihood of a household to complete the interview over the phone. These results informed updates to parameters (i.e. maximum number of call attempts) in a way that reduces the overall number of contact attempts made while minimizing the impact on response rates. We also analyzed how to optimize CATI interviewer work schedules to increase staffing when contact attempts are most effective. Together, these results utilize existing data about respondents and previous contact attempts to develop more efficient and targeted CATI contact strategies.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

An Experiment with Varying Survey Cover Letter Paper Color and Its Effect on Survey Response Rates by Mode

Andrea H Schwanz *IMPAQ International*

Amy LDjangali *IMPAQ International, Inc.*

Michelle A Cantave *IMPAQ International*

One tool employed in the effort to increase response rates to mailed surveys is colored paper. The color may make the materials stand out among other mail, leading to more recipients noticing the mailing and subsequently completing the questionnaire. Results are mixed, however, with some studies showing little or no effect of questionnaire paper color while another meta-analysis found a significant effect. IMPAQ conducted a randomized experiment to investigate the effects of paper color for a statewide employer survey. Unlike previous research, we varied the color of the cover letter, not the color of the questionnaire. The experiment was conducted as part of a multimode survey of employers for an evaluation of a state workforce program. Employers (N = 5,893) were randomly divided into four groups. There were two mailings sent out, an initial mailing with a copy of the questionnaire and a Web survey link, and a second mailing about a month later with a reminder and another copy of the questionnaire and the Web survey link. Half of the sample received white cover letters for both mailings (white/white), and the other half were evenly divided between three groups – blue/blue, white/blue, or blue/white. Our analyses will examine the effect of the cover letter color on overall response rate, wave response, response mode, and length of time from mailing to response.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

How We Ask About Gender

Kalina Popova *Nielsen*
Kathy Steinberg *Nielsen*

This project will perform a test of alternative ways of asking survey respondents about their gender. Heretofore the gender options were a simple Male/Female dichotomy, without any other categories or a 'Decline to Answer' option. Based on recent industry practices, client concerns and discussions on industry forums (namely AAPORnet) we expect to test at least three different ways of asking about gender with the end goal of evaluating the best alternative to use going forward, both in terms of respondent experience and data validity. Respondent experience will be evaluated based on suspend rates during field, non-response to the question, length of interview, and other indicators of the respondent experience. Once we have collected the data we will tabulate the responses we received with the different question specifications, correlate those with the gender information we have stored in the respondent's panel profile, and analyze the relationship among the alternative ways of asking gender. In addition, we will evaluate the relationship between alternative gender specifications and the way we ask about sexual orientation (LGBQ - Lesbian, Gay, Bisexual, Queer) status. The research will have a base of 2000 gen pop respondents and oversamples for Hispanic, Asian, and African American respondents to allow for sufficient bases for subgroup analysis. In addition to demographics and survey experience questions we will also include 5 to 10 benchmark questions which we will then use to evaluate whether the alternative gender questions result in different levels of data validity (i.e. closeness to known benchmarks).

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Designing a Household Survey of Occupational Injuries and Illnesses

Lisa Lee *NORC, Univ. of Chicago*

Kristen Neishi *NORC*

David Gleicher *NORC*

Robin Kaplan *Bureau of Labor Statistics*

Nola du Toit *NORC*

Kennon Copeland *NORC*

The Survey of Occupational Injuries and Illnesses (SOII) is the primary source of information for the Bureau of Labor Statistics (BLS) on nonfatal workplace injuries and illnesses. In this surveillance system, sampled employers maintain logs of workplace injuries and illnesses that occur to workers during the calendar year that meet recordability requirements established by the Occupational Safety and Health Administration. The SOII estimates are subject to underreporting due to potential filtering effects of injury and illness reporting that can lead both employers and workers to not report in-scope injuries or illnesses. For example, employers may wish to underreport or misclassify injuries and illnesses to keep injury and illness rates low. Further, the employer may not be aware of reportable events that occurred because some workers do not report incidents to avoid losing their job or other opportunities such as a raise or promotion. One way to reduce underreporting due to filtering effects associated with employer reporting of injuries and illnesses is to collect this information directly from workers in a household survey. NORC assisted BLS in developing a sample design and questionnaire for a household survey on occupational injuries and illnesses (HSOII). We present the work that NORC completed in designing a questionnaire for the HSOII. The questionnaire was tested and refined in more than 60 interviews over three rounds of cognitive testing. We address the challenges in translating the establishment SOII forms into a household survey questionnaire and potential risks to data quality in the household survey format, such as error due to respondent comprehension and recall issues as well as those that result from proxy reporting.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Long vs. Wide Question Format?

Benjamin J Earnhart *ACT, Inc.*

Our organization was administering questions with a wide format on a series of conference evaluation surveys. We decided to try a long format. First, for one conference, participants were randomly assigned to long vs. wide. Based on our readings and observations, we derived the following hypotheses: 1) Moving to the new format will reduce item non-response. 2) The frequency of comments to open-ended questions should increase. 3) The length of comments should increase. 4) Smartphone users are most likely to be affected by the change, potentially in all four of the ways listed above. In the experimental data, item non-response was found to decrease, and the length of comments increased. Based on the promising results of the experiment, we switched entirely to the long format. With the increased statistical power of combining results from several conferences, all four hypotheses were strongly supported. A small change to our survey layout improved the quantity and quality of the data we gather.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Identifying the best times for cognitive functioning using survey methods: Matching university times to student chronotypes

Paul Kelley *The Open University*
M.D.R. Evans *University of Nevada*
Jonathan Kelley *University of Nevada*

Education and work generally have fixed starting times that are not adjusted for different chronotypes. In education later starting times are recognized as crucial to student sleep, health and performance. This study used two new approaches to determine start times that align undergraduates to optimal times for cognitive performance. The first is a survey-based, empirical model, and the second a neuroscience-based, theoretical model. The survey model focused on individuals' self-reported chronotype and best times for their university work. Using this approach data from 190 first and second year university students was collected. This data was analyzed to determine optimal times for cognitive performance. The neuroscience-based theoretical model involved creating a generalized solution to determine working hours based on research in sleep, circadian neuroscience, fMRI circadian research, and practical considerations. Strikingly all three of these models are broadly compatible in indicating later start times, but the survey and neuroscience models add several important points: 1) While confirming the value of later starts, they indicate that much later starting times (after 11:00) are optimal; 2) There is no single starting time that does not disadvantage one or more chronotypes; and 3) The best practical model may involve three alternative starting times with one early afternoon shared session. The implications are briefly considered. The survey questionnaire can be readily adapted to other populations.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Living up to Your Potential: Convincing Field Staff to Consider Modeled Propensity

Ryan Hubbard *Westat*

Westat developed a propensity model for a large-scale, national longitudinal study intended to improve efficiency by identifying pending cases likely to result in a completion at that specific week in the field period. This dynamic model incorporates current and previous round case history and record of call information and is updated weekly based on the specific conditions for pending cases at that time. The model is useful for identifying higher potential cases (likely to result in an appointment or completion the next attempt), enabling supervisors to direct interviewer effort toward these cases and away from cases not likely to result in completion; however, there have been a number of challenges to integrating this model into daily operational field tasks. Over the span of four rounds of data collection, the project implemented a set of evolving procedures in an attempt to gain buy-in from field management staff, integrate the model information into reports and eventually into the electronic field management system, and provide feedback on the use of the model to the field. A number of steps were implemented to adjust the timing and output of the model and even place certain cases on hold to better fit the model to the realities of the operational environment. Results detail the efficacy of the model implementation over time and the results of operational adjustments designed to overcome the largest obstacle – field acceptance and integration of the potential rating procedures.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Response order effects for item on educational attainment

Salima Douhou *City, University of London*

The use of show cards is common for closed-end items in face-to-face surveys. The European Social Survey includes an item on educational attainment and the Showcard for this item contains a list with many items that respondents might be less likely to read out completely during the interview. Several countries within the ESS use labels while others do not use labels on the Showcard of the educational attainment item. There are also countries that switched from no labels to using labels between rounds of the ESS. From a cross-national research perspective, it would be very relevant to find out what we should recommend: labels or no labels on these type of show cards. We hypothesize that the cognitively challenging task of choosing a response option from a list with many items and long labels can be simplified by labelling the items. We find that response time is significantly reduced when labels are used and it is less obvious to complexity of the item and showcard are related to response order effect due to the nature of the item.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Mailing to Millennials: How to Reach Young Adults with Mail-Based Surveys

Yelena Pens Moore *Nielsen*

Austin Countryman *Nielsen*

Robin Gentry *The Nielsen Company*

In the 21st century, young adult participation has been a continued concern for survey research organizations, including Nielsen, a leading global information and measurement company. Historically, young adults, particularly in the 18-24 demographic, have been a hard-to-reach population due to the difficulty of making contact with this population and low response rates once contact has been made. Qualitative research techniques, in particular focus groups, have become standard methodology for revealing a variety of information and providing insight into specific research groups. Nielsen conducted focus groups in fall 2016 with 18-24 year olds, particularly with young adults who live with only other young adults. The overall goal of the research was to better understand this population as well as to find actionable insights in order to increase response rates and representation of young adults in the Nielsen radio diary service. The main focus of the study included ideal methods to reach 18-24 year olds via a mail-based survey and then to engage them in the survey process. Additionally, findings included reasons for attrition with multiple surveys as well as optimal incentive strategies. Finally, the results included a comparison of diverse groups of 18-24 year olds from previous survey respondents to general population. In this presentation, we will answer the research questions: (1) what are the motivators for young adults to participate in a survey, and (2) what is the optimal method to reach and recruit young adults in a mail-based survey?

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

Measuring Voter Apathy in Georgia: Survey Design Initial Testing and Validation

John R. Barner *Carl Vinson Institute of Government, University of Georgia*

Michelle Bailey *Carl Vinson Institute of Government, University of Georgia*

Theresa A. Wright *Carl Vinson Institute of Government, University of Georgia*

As Dean (1960) noted, political apathy has “generally been defined simply as voting or nonvoting; most of the generalizations which have been advanced have been based on studies using this single criterion (p. 187). However, recent research studies have advanced a view that interactions between several indicators result in a behavioral shift from voting to nonvoting (Groenendyk, 2016; Pratchett, 1999; Teixeira & Halpin, 2010). This paper presents a new multi-indicator measure of political apathy and describes its development, testing, validation and initial findings. While political apathy as measured by voting/nonvoting behavior reduces to two dimensions, we argue that voters routinely engage with indicators that then influence eventual voting behavior. This multi-indicator measure, the Brief Political Apathy Survey (BPAS) provides responses for identified factors including levels of political involvement outside of voting, engagement with political media, real or perceived barriers to political activity and/or voting and a “milieu effect” of interpersonal relationships and circumstances that may influence respondent’s attitudes toward the political process and level of political apathy. The initial instrument was developed and pilot tested in a statewide poll with voters in Georgia. After submitting the data to factor analysis, adjusting the instrument and assessing the reliability and validity of each of the instrument domains, the resulting instrument is robust and tailorable to national surveys of political apathy, voter alienation and voter bias. These future applications of this measure show great promise for determining a data-driven profile and substantive explanation of how and which events move the mood of voters as a whole or in part.

May 18th, 2017
3:30 PM - 4:30 PM
Concurrent Session P

Poster Session #1

The rise of authoritarianism in America: contrast of electoral behavior in US and Mexico

Diana Paola Penagos Vasquez *Parametria S.A. de C.V.*

Francisco Abundis Luna *Parametria S.A. de C.V.*

Jose Alberto Vera Mendoza *Parametria S.A. de C.V.*

The results of the recent presidential elections in the United States were surprising because of the two controversial candidates in contest. From Nazi Germany, political science has studied the authoritarian political figures of greater salience in certain historical contexts. The literature has shown that authoritarian leaders represent the fulfillment of extreme policies for keeping the order and enforce the law. Different researchers emphasize the importance of studying the psychological profile of those who support them. What factors lead voters to support this type of political figures? Karen Stenner (2005), Hetherington and Weiler (2009), Matthew C. MacWilliams (2016) have proved that there is a group of people with authoritarian tendencies that can be mobilized in times of social instability or perception of physical threat. The authors show that despite there are authoritarian voters across all the American politics spectrum, recent trends indicate that they are increasingly aligned with the Republican Party and Trump is the leader who gives voice to these voters. Smith and Cohen (2016) state that "Latin American elections hold clues to understanding who supports authoritarian leaders". This research sought to verify if this type of voter profile exists in Mexico and which political figure they would support. Using a new national survey of Mexican voters, this study reproduced the Matthew C. MacWilliams' work. The results show that there is an important part of the Mexican electorate that is authoritarian, and the electorate exhibit statistically significant and substantive authoritarian attitudes on different issues. Authoritarianism is an important predictor of people's likelihood of voting for Lopez Obrador in Mexico. According to Smith and Cohen's results, the education level is other predictor for Obrador's voters. The results show that the authoritarianism is a phenomenon that broke into American public view but will persist long for the entire continent.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

The State of the Glass Ceiling in the Wake of Hillary Clinton's Candidacy (Panel)

The State of the Glass Ceiling in the Wake of Hillary Clinton's Candidacy

Marjorie Connelly *AP-NORC Center for Public Affairs Research*

Emily Swanson

Emily Swanson *The Associated Press*

Jennifer DePinto *CBS News*

Dan Malato *AP-NORC Center for Public Affairs Research*

Marjorie Connelly *AP-NORC Center for Public Affairs Research*

Jennifer Benz *NORC at the University of Chicago*

Trevor Thompson *NORC at the University of Chicago*

Jocelyn Kiley *Pew Research Ctr.*

Ninety-six years after women were granted the right to vote in the United States, Hillary Clinton was the first woman nominated for president by a major-party. In an Associated Press-NORC Center for Public Affairs Research poll, 3 in 10 Americans expected Mrs. Clinton's gender to hurt her electoral chances, even as three-quarters said women and men are equally good political leaders. How did Mrs. Clinton secure the Democratic Party's nomination? With exit poll data from the Democratic primaries, Emily Swanson will demonstrate the how Mrs. Clinton was able to avert the challenge from Sen. Bernie Sanders. And, by analyzing general election exit poll numbers, Emily will illustrate gender differences that impacted the 2016 election. Even as Mrs. Clinton was making history as the first major party presidential candidate, how much gender discrimination do women continue to face? Jennifer DePinto from CBS News will provide results from a CBS News/New York Times poll which indicates that while women feel positive about the advances that have been made and optimistic about the future, they recognize that challenges still exist. Data from The AP-NORC Center poll presented by Marjorie Connelly will examine the public's views about women in politics. And gender differences on the level of difficulty women continue to face will be the focus of the paper based on polls conducted the Pew Research Center and presented by Jocelyn Kiley.

Gender in 2016: A View from the Exit Polls

Emily Swanson

With the first-ever female presidential candidate on the ballot, there was a historically large gender gap in the 2016 United States presidential election. But contrary to many expectations, it wasn't because of women flocking in record numbers to support Hillary Clinton's candidacy. According to exit polls conducted by Edison Research for the Associated Press and television networks, most women did support Clinton, but Trump's winning margin among men was equally large. Crucially, Trump was supported by most white women, who have generally supported Republican candidates in the last several presidential elections, and drew particularly strong support from women in his white working class base. He equaled Clinton's support among married women and those ages 45 and over, and prevented defections among Republican women. Women were more likely than men to say they were significantly bothered by Trump's behavior towards women, but among both men and women only those who said they were most bothered by Trump's behavior supported Clinton, with those saying they were only somewhat bothered or less breaking strongly toward the Republican candidate. Meanwhile,

Trump improved over 2012 margins among some groups of men. This analysis explores gender dynamics in the 2016 election using data from exit polls conducted at precincts across the country and through absentee phone interviews, highlighting how support by gender did and didn't shift since earlier elections.

Women are Optimistic but Challenges Remain

Jennifer De Pinto

2016 represented the first time in U.S. history that a woman was a presidential nominee of a major political party. So, have women come a long way? CBS News and The New York Times conducted a nationwide survey to ask women about their opportunities to succeed in life and how they feel about the future of girls in the U.S. The survey explored women's views (as well as men's) of sexism, sexual harassment in the workplace as well as equal pay. These results will be presented along with women's views on Hillary Clinton's run for the presidency and how much her gender had an impact.

Hillary Clinton Was the Democratic Nominee for President. Now What?

Dan Malato

Although a woman was a major party standard bearer, a national poll conducted last August by The Associated Press-NORC Center for Public Affairs Research found most Americans think women continue to face obstacles in politics. Large majorities of Americans said they regard women political leaders as equal to men and think a woman would be up to the challenges a president may face. Yet, about half of the public said women still have fewer opportunities in politics than men. Only 3 in 10 Americans expected Clinton's gender to hurt her chances on Election Day. The same number said she was being held to a higher standard than other candidates because she is a woman. Women were particularly likely to say they are given fewer chances to succeed in politics. They were also more inclined to see Clinton's gender as a disadvantage, while men tended to say the fact that she is a woman would help her chances of being elected. Still, 7 in 10 said in August that the historic nature of Clinton's candidacy would have no bearing on their own vote choice in November. Nearly 20 percent said the opportunity to elect the first woman president would make them more inclined to vote for Clinton, and about 10 percent said it would make them less likely to vote for her.

Gender attitudes and the 2016 election

Jocelyn Kiley

Throughout the 2016 election, a frequent question was how much Hillary Clinton's gender "mattered" in the election. In particular, whether it was a liability. Most voters, if asked, said Clinton's gender mattered little to their vote choice. And in a Pew Research Center survey from June 2016, 40% of voters expected that the fact that she was a woman would help her with voters in the election, while 45% said it would make little difference (just 12% said it would hurt her). This paper will attempt to shed more light on this question by examining the relationship between gender attitudes, vote choice and views of the candidates in 2016. Primarily through the use of Pew Research Center's American Trends Panel (a nationally-representative, probability-based online panel survey with more than 5,000 panelists), the paper will look at views of Clinton, Trump and other aspects of the 2016 campaign over the course of the year – this includes during the primaries, the general election, and post-election. It will explore how these views were associated with measures of general attitudes about gender equality and roles (e.g., views about women's opportunities and gender roles in relationships and families), not only among all voters, but also within different subgroups (by gender, education and party, in particular). It will also

examine the relationship between gender attitudes and other key political attitudes, including views of race, immigration and social values.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Practical Advice for Mail Mode Surveys (Panel)

Practical Advice for Mail-Mode Surveys

Kate Williams *Nielsen*
Robin Gentry, Kate Williams
Robin Gentry *The Nielsen Company*
Yvonne Shands *SSRS*
Don Dillman *Washington State University*
Pierce Greenberg *Washington State University*
Elizabeth Poehler *Census Bureau*
Dorothy Barth *Census Bureau*
Michael Stern *NORC*
Ipek Bilgen *NORC*
Ned English *NORC*
Erin Fordyce *NORC*

We are all too aware of the current state of telephone interviewing—declining number of landlines, TCPA concerns, and increasing cost. In response, we have seen a resurgence of mail-mode surveys. Mailing, while it avoids certain issues, presents challenges of its own. In this era of rapid change it is not only critical that researchers understand how to conduct a methodologically sound mail-mode survey, but it is equally important that researchers are able to reach increasingly diverse population through the mail. This panel session focuses on current, actionable best practices in mail-mode surveys to achieve improved response rates and sample representation. A diverse panel of experienced researchers will discuss the effectiveness of various outer envelopes/ mailing packages, use of pre-alerts and a discussion of their effectiveness, impact of addressing a package by respondent name or a generic salutation, offering multiple modes of response and/or sequencing of modes, and the nature of communication attempts made with sample members.

Mailing Methods Matter

Robin Gentry, Kate Williams

Researchers are aware of the myriad of choices to be made when designing a mail-mode survey. Each methodological decision can alter potential respondents' behavior, and when several such decisions are considered cumulatively, they can impact the success of an investigation. We currently work in the context of declining response rates, which makes every detail matter more. This presentation focuses on methodological choices related to the mailed materials themselves, which can have a significant impact on people's willingness to cooperate with a survey request. The data for these tests are based on a series of experiments conducted in several large- and medium-sized U.S. cities across a five year period. Sample members were randomly assigned to treatment and control groups. Treatment groups included options for sample pre-screening, the addition of pre-alerts, the addition of a reminder postcard, variation in how the envelope is addressed, and variation in the envelope itself. We will present the outcome of the experiments and the impact those design choices had on survey participation. When possible, we will provide advice about best practices for mail-mode surveys and in some cases, we will suggest areas where more research should be considered.

Operational Lessons Learned with Mail-Mode Surveys

Yvonne Shands

With the increasing popularity of mail-mode surveys, it behooves survey professionals to educate themselves on how to ensure their mail-mode surveys are conducted efficiently. For the past 8 years, SSRS has supported the mailing, scanning, and data entry for a Nielsen mail-mode survey. This paper focuses on operational lessons learned over these years and how to measure and weigh operational efficiency with methodological rigor as well as fiscal responsibility. Challenges and lessons learned to be included in this panel section include: size, fit and fold patterns for mail surveys as they relate to their outbound as well as Business Reply envelopes; impact of type of jogging and opening machinery used; print processes and how they interact with weather elements in the mail stream; effective use of 2D barcodes and Intelligent Mail Barcodes (IMBs); the role of usability testing in designing Post Office undeliverable return receipting systems; effect of respondent choice of writing implements as well as researcher choice of coloration and format on how well surveys can be scanned; and best practices in assembling a mail piece.

What Makes Mail Survey Requests Effective: Results from an Experiment

Don A. Dillman, Pierce Greenberg

One of the most understudied factors that may affect mail survey response is the nature of the communication attempts made with sample members. In this presentation we will report results from an experiment that evaluates a comprehensive and integrated design of all communication components, the design of which is based in part on the “pre-suasion” theoretical perspective offered by Robert Cialdini. Results of this approach are compared against a standardized approach that incorporates common government survey practices across multiple contacts. The communication process is conceptualized as including all contents of these contacts, including such considerations as the outside of the mail envelope, contents of the envelope, design of the questionnaire cover pages, and initial questionnaire content. The experiment uses address-based sampling of households in a lower internet penetration area of the United States. Our goal is to provide insight into how communications with respondents affect their response behavior and might be improved.

Influencing Response Mode Choices in the American Community Survey

Elizabeth Poehler, Dorothy Barth

In 2013, the American Community Survey (ACS) added Internet as a self-response option, in addition to the existing paper questionnaire. While early research indicated that offering the Internet as a response mode during mail data collection decreased response rates, research conducted in 2011 indicated that response rates could be improved with an Internet response option if offering a paper response option was delayed. Research conducted in 2014 and 2015 evaluated the usefulness of a pre-notice letter in this context, first looking at the impact of simply dropping the pre-notice letter and then later evaluating the impact of both dropping the pre-notice letter and mailing the Internet invitation earlier. Additional research in 2015 focused on the language in the later mailings to reduce the emphasis on choice of response mode. Current efforts are underway to test a more targeted mailing strategy that would include a “Choice Strategy” in the initial mailings for certain areas of the country. This paper will provide information on the overall ACS mail strategy, findings from our 2014 and 2015 studies about the usefulness of pre-notice letters and language around choice, and our plans for future testing of a targeted mailing strategy.

Understanding the Strengths and Roles of the Mail Mode in Sequential Multi-Mode Surveys: Evidence across Recent Experimental Studies

Michael J. Stern, Ipek Bilgen, Ned English, Erin Fordyce

Despite the wide scale use of mail surveys, researchers have consistently expressed two concerns. First, mail is considered a “higher cost” mode of self-administered data collection in comparison to the web (Dillman et al. 2009) due to the expenditures associated with postage, processing, and printing (Groves et al. 2009). Second, mail surveys are seen as a less flexible mode given it is difficult to use branching or probe questions as effectively as in interviews or on the web, thus limiting the complexity and in some cases depth of a survey instrument (de Leeuw 2009). However, mail surveys have continued to do reasonably well in terms of response especially when used in conjunction with another form of data collection such as web (Stern, Bilgen and Dillman, 2013). As a result of both the limitations and strengths of this mode in mind, this paper explores data from five different experimental, sequential multimode studies, all of which are different in some an important way, to examine how mail performed relative to web and telephone as part of multimode, sequential design. Specifically, the data we analyze cover response rates across modes, substantive response differences, geography of response by mode, demographics difference in response by mode, and the variation in questionnaires as part of our analyses. We then compare the findings to an experimental study using three modes—mail, web, and telephone—fielded concurrently to assess how findings from the multimode approach impact our findings. In doing so, the paper ultimately furthers our understanding of the strengths and weaknesses of the mail mode as part of a sequential multimode design and in isolation.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Advances in Measuring Sexual Orientation and Gender Identity (Panel)

Advances in Measuring Sexual Orientation and Gender Identity

Philip Brenner *University of Massachusetts Boston*

Stephanie Marken, Evan E. Krueger, Ilan H. Meyer, Walter O. Bockting, Sari L. Reisner, Jody L. Herman

Matt Jans *UCLA Center for Health Policy*

Stephanie Marken *Gallup*

Ilan Meyer *Williams Institute, UCLA School of Law*

Evan Krueger *Williams Institute, UCLA School of Law*

Walter Bockting *Columbia University Medical Center*

Sari Reisner *Harvard Medical School*

Jody Herman *Williams Institute, UCLA School of Law*

Nancy Bates *U.S. Census Bureau*

Jennifer Ortman *US Census Bureau*

Erin Fordyce *NORC at the University of Chicago*

Michael Stern *NORC at the University of Chicago*

Melissa Heim Viox *NORC at the University of Chicago*

Ipek Bilgen *NORC at the University of Chicago*

Sabrina Bauroth *NORC at the University of Chicago*

Stuart Michaels *NORC at the University of Chicago*

Christopher Harper *CDC Division of Adolescent and School Health*

Michelle Johns *CDC Division of Adolescent and School Health*

Richard Dunville *CDC Division of Adolescent and School Health*

Justine Bulgar-Medina *University of Massachusetts at Boston*

Change in the U.S. social and political environment has brought mainstream attention to sexual minorities. As a consequence, major surveys have begun or expanded their measurement of sexual orientation and gender identity to meet the growing demand for data on these populations. Thus, evaluating and implementing methods of measuring gender identity and sexual orientation is more important than ever. The major challenge is to do this in a manner that adequately and accurately represents all respondents' identities and also provides useful statistics on these populations. To these ends, this panel will discuss current and developing research on measurement of sexual minority identities, sexual/gender identity, and sexual orientation. Panelists will discuss findings from national and state production surveys, survey experiments, and question evaluations and pilot studies. Papers focus on evaluating data quality at various stages of the survey process, including respondent recruitment and question evaluation, using a wide variety of frames and samples, including household, web panel, racially diverse, and adolescent and young adult samples. Beyond assessing the current status of survey research on sexual minorities, panelists will note areas of the greatest need for future research.

Sensitivity and Specificity of a 1-item Assessment of LGBT Identity in a National Sample of U.S. Adults

Stephanie Marken, Evan E. Krueger, Ilan H. Meyer, Walter O. Bockting, Sari L. Reisner, Jody L. Herman

An important first step in measuring issues of importance to the lesbian, gay, bisexual and transgender (LGBT) population is to accurately identify individuals' gender identities (GI) and sexual orientations (SO). Survey researchers have experimented with several methods of identification to determine which approach best covers the LGBT population while minimizing respondent burden. We tested two approaches: a one-question assessment of LGBT identity versus 3 questions to measure SO and GI. In the one-item approach, respondents are asked if they identify as lesbian, gay, bisexual or transgender, providing a "yes" or "no" response. The 3-items approach assessed SO and GI separately. To assess SO respondents were asked "which of the following do you consider yourself to be?" Respondents select from multiple response options, including "straight or heterosexual," "lesbian," "gay," "bisexual," "queer," and "same-gender loving." To assess GI, respondents were first asked what sex they were assigned on their original birth certificate. Then, they were asked to choose the label that describes their current GI, with the response categories, "woman," "man," "transwoman," "transman," and "non-binary/gender queer." The one-item approach has obvious economic advantages as it provides a more efficient data collection. However, this approach may not be as precise and may lead to undercoverage of the population. To measure the sensitivity and specificity of the one-step assessment, Gallup worked with researchers at UCLA and other universities to compare the two approaches using the Gallup Daily Tracking survey, a nationally representative study of all U.S. adults, aged 18 and over, conducted using a dual-frame, random-digit-dialing (RDD) sampling frame. We present differences between the two approaches based on data from approximately 7,700 respondents. We will present findings from this research, including interviewers' experiences implementing each approach and the accuracy associated with covering the population using the different methods.

Comparing Two Versions of a 2-step Assessment for Identifying Transgender Respondents in a National Sample of U.S. Adults

Ilan H. Meyer, Even E. Krueger, Stephanie Marken, Sari L. Reisner, Walter O. Bockting, Jody L. Herman

In recent years, transgender population surveys have gained interest from researchers and pollsters. A 2-step measure is most often recommended where respondents report their sex assigned at birth (male or female) and their current gender identity (man, woman, or transgender). Respondents are classified as transgender if their current gender identity is "transgender" or if their current gender identity is different than their reported sex assigned at birth. We tested two versions of the 2-step question method that differ in the way the sex assigned at birth question is asked—version 1 leaving an open question and version 2 providing the "male" and "female" options in the question. The versions also differ in the response options regarding gender identity—version 2 including only "transgender" in addition to "man" and "woman"; version 1 including more response options (e.g., "genderqueer"). We compared these in a study of Gallup Daily Tracking survey, a nationally representative study of U.S. adults, aged 18 and over, conducted using a dual-frame, random-digit-dialing (RDD) sampling frame. Version 2 was more efficient in reducing refusal and Don't Know answers from 7% in version 1 to 0.1% in version 2. Version 1 yielded more respondents who identified as transgender than version 2 (0.53% vs. 0.37%). Compared to version 1, version 2 was not efficient at identifying non-binary/genderqueer individuals (0.4% vs. 0.06%). We also assessed whether the 2-step method of asking "sex assigned at birth" would be a good replacement for a standard Gallup assessment of sex. We found that against the standard sex question, both versions of the sex assigned at birth question performed extremely well, with sensitivity of over 99% for each.

Collecting Sexual Orientation and Gender Identity with a Household Proxy

Nancy Bates, Jennifer Ortman

Recently, a great deal of attention has been paid to data collection on sexual minorities. In 2011, the National Institutes of Medicine recommended that sexual orientation and gender identity (SOGI) be added to federally-funded surveys and in 2014, the Office of Management and Budget formulated an interagency workgroup on improving measurement of SOGI in federal surveys. One topic garnering attention is whether SOGI can be successfully collected via proxy. To acquire empirical evidence on the subject, the Census Bureau sponsored the 2015-16 Joint Program in Survey Methodology Practicum. The primary research goal was to collect SOGI for all members of a household aged 16 and older using a single household informant. This methodology mimics how demographics are collected for surveys such as the Current Population Survey and American Community Survey. The Practicum class designed an online survey administered using Amazon Mechanical Turk. The survey was completed by a single household member providing an opportunity to evaluate the performance of questions in a proxy environment. The survey consisted of two stages, a “cognitive” pre-test and a main survey. The cognitive survey collected self-reported and proxy data on SOGI and included follow-up questions about sensitivity, confidence, and comfort levels when reporting for others. The main survey collected data on SOGI, employment status, and income, but also contained a split panel gender identity wording experiment. Data was collected from 5,500 total respondents. Research Questions: What is the rate of nonresponse to SOGI questions among those who answer on their own behalf? Compared to standard demographics? Compared to historically sensitive questions? What is the rate of nonresponse to SOGI questions when asked by proxy? How confident are proxy reporters in their SOGI answers? How comfortable are they when reporting for others? Which question wording yielded the lowest measurement error?

Developing Effective Methods and Social Media Recruitment for Adolescent Sexual Minority Males and Transgender Youth: An Examination of Platforms for Recruitment, Demographics, Targeting, Costs, and Overall Feasibility.

Erin Fordyce, M Stern, M Heim Voix, I Bilgen, S Bauroth, S Michaels, C Harper, M Johns, R Dunville

Researchers face many challenges when recruiting and surveying LGBTQ youth including a lack of disclosure to others and an evolving sexual identity. Our study collects data from two of the least studied segments of the LGBTQ population, adolescent sexual minority males age 13 to 18 years old and transgender youth age 13 to 24 year old. Our presentation reports findings from a pilot study that used social media ad campaigns to recruit these groups for a web survey on HIV prevention. Recruitment ads were designed for young gay, bisexual, and questioning males and transgender youth. Ads were posted to six social media sites—Facebook, Snapchat, Instagram, Twitter, Kik, and Tumblr. Multiple social media sites were used to ensure sufficient representation of the age ranges and ample representation of Black and Latino youth who are at especially high risk for HIV infection. We discuss demographics for eligible respondents screened into the survey versus those respondents completing the survey. We further explore the costs and feasibility of recruiting these hard-to-reach populations and the performance of each social media site in recruitment efforts. Finally, we compare the recruitment success of targeted versus specific ads. Results will aid our understanding of using various social media sites for survey recruitment as well as targeted ads for survey recruitment.

Race and Sexual Orientation: Cognitively Testing Expanded Categories for Sexual Orientation with Racial Minorities

Justine Bulgar-Medina

There has been a steady increase in research on sexual minorities (i.e. lesbian, gay, bisexual or transgender (LGBT)) on topics ranging from demographics to health disparities (Black et al. 2000; Gates

2014). As research attention for this population increases, the need for reliable measurement of sexual orientation becomes ever more important. Previous research suggests that identity development and disclosure practices are not constant over the life course and most notably do not report on the cognitive processes a respondent goes through during recall (Calzo et al. 2011). Beyond these general concerns surrounding recall of identity development and disclosure, the disclosure and identity management practices for multiple minorities (racial and sexual) is likely substantially different when understood in context (Goode-Cross and Good 2009). This work presents the findings from nearly two dozen cognitive interviews conducted on sexual orientation self-identification, behavior, attraction and disclosure with a specific emphasis on the contextualized responses of dual (racial and sexual) minorities. Two core sets of questions were tested: (1) the expanded categories of sexual orientation (e.g. including queer as a response category); and (2) the practices of disclosure of identity in social contexts (e.g. the workplace) using established measures (Lidderdale et al. (2007) and Mohr & Kissinger (2000)). Participants were recruited through community groups and centers via flyers and key informants. Cognitive interviews were conducted face-to-face (if local to the researcher) or via video/online chat (if not local or preferable for respondent).

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Enhancing Federal Statistics Using New Data Sources While Respecting Privacy (Panel)

Enhancing Federal Statistics Using New Data Sources While Respecting Privacy

Brian Harris-Kojetin *National Academy of Sciences*

Brian Harris-Kojetin

Frauke Kreuter *Univ. of Maryland, JPSM*

Colm O'Muircheartaigh *Univ of Chicago NORC*

Michael Link *Abt SRBI*

John Eltinge *U.S. Bureau of Labor Statistics*

Federal statistics provide vital indicators of wellbeing of the population. For decades the sample survey has been a primary method of collecting data for federal statistics. However, the costs of conducting such surveys have been increasing while response rates have been declining, and many surveys are not keeping up with growing demands for more timely and detailed local information, rather than national information. The Committee on National Statistics, a standing board at the National Academy of Sciences, Engineering, and Medicine convened a committee of experts in social science research, sociology, survey methodology, economics, statistics, privacy, public policy, and computer science to explore a possible shift in federal statistical programs—from the current approach of providing users with the output from a single census, survey, or administrative records source, to an approach combining data sources to give users richer and more reliable datasets. This AAPOR panel will feature presentations and discussion about the first report from this National Academies' panel. The report reviews the current approach for producing federal statistics, examines other data sources that could also be used for federal statistics, and discusses the environment needed for utilizing multiple data sources in the future, including statistical methods of combining data sources, mechanisms for research access, and approaches for protecting privacy and preserving confidentiality. There will be three presentations by the study director and two panel members covering the content of the report. There will also be two discussants, who will offer perspectives on the report from the federal statistical community (John Eltinge) and the private sector (Michael Link). Please note that the report is currently in review at the National Academies, and we are not permitted to reveal conclusions or recommendations from the report until its publication, anticipated in early 2017.

Current Challenges and Opportunities for Federal Statistics

Brian Harris-Kojetin

Statistics for the common good are embedded in the very foundation of the United States and are central for our democracy and economic and social wellbeing. Sample surveys have been a primary method for collecting data for federal statistics for decades. However, continued reliance on sample surveys as a principal means of collecting national statistical data are threatened by the increasing difficulty and cost in conducting these surveys with consequent threats to data quality, and by the increasing demand for more geographically detailed and faster information. This presentation will discuss the importance of federal statistics for the country, the foundational principles for federal statistics, and the current environment for conducting federal surveys. It will provide an overview of the panel's work and set the context for the other presentations. The National Academies panel was funded

by the Laura and John Arnold Foundation to explore a possible shift in federal statistical programs—from the current approach of providing users with the output from a single census, survey, or administrative records source, to a new paradigm of combining data sources with state-of-the art methods to give users richer and more reliable datasets that lead to new insights about policy and socioeconomic behavior. In carrying out its work, the panel sponsored three public workshops examining benefits and challenges with the use of administrative and private sector data sources by federal statistical agencies, state and local governments, and private sector firms. Topics included integrating multiple data sources and preserving privacy. A discussion was also held with the heads of the thirteen principal statistical agencies as part of the panel's fact-gathering activities.

Using Administrative and Private Sector Data for Federal Statistics

Frauke Kreuter

In recent years we have witnessed an explosion of data coming from many sources, including streaming data production (e.g., utility meters, traffic cameras, and other sensors), internet behavior documentation (e.g., browser search terms), and social media posting (e.g., Twitter, Facebook, LinkedIn data) as well as consumer information data (e.g., Zillow, Experian, and other credit bureau data). Some of these new data resources arise from digital records of government agencies (e.g., the health care transaction records of the Centers for Medicare and Medicaid Services). But many of these sources lie in private sector enterprises. Indeed, a whole set of new enterprises are using large digital data resources as the prime engine underlying their business model (e.g., Uber, AirBnB, LinkedIn). There have also been several initiatives to improve evidence-based policymaking, which have emphasized the importance of reusing existing administrative data that the government already has available. Government administrative data i.e., data collected by government entities for program administration, regulatory, or law enforcement purpose, such as federal tax forms, can be a valuable source of information for federal statistics because they often provide individual level data on an entire population that would be cost prohibitive to obtain via a survey. In this presentation we will discuss the different kinds of data from federal administrative and private sector sources and how the characteristics of these data affect their potential utility and usability for federal statistics. We briefly review efforts by national statistical offices around the world to examine and experiment with using these data sources to produce official statistics. We then review current work in the U.S. to examine and evaluate these new data sources for federal statistics. We discuss the benefits and challenges of using these sources for statistical purposes.

Combining Data Sources While Protecting Privacy

Colm O'Muircheartaigh

Although federal statistical agencies provide descriptive statistics that are policy relevant, data from the federal statistical agencies are also a key resource for policy analysis and evaluation conducted by the research and evaluation communities. Indeed, for decades US society has profited from applied social science research and policy analysis using federal data, conducted by universities, non-profits, think tanks, and advocacy groups, supported by government grants and contracts, private foundations, and corporate and individual donors. In this presentation, we briefly review the legal, policy, technological, and procedural issues associated with protecting privacy while permitting access to federal data for research purposes, including current challenges statistical agencies face when providing data access and statistics in an ever-changing data environment. We summarize a variety of approaches that federal statistical agencies employ for providing access to confidential data for statistical purposes as well as access models from other countries, focusing on those that include combining multiple data sources. We include a brief discussion of some relevant Privacy Enhancing Technologies ("PETs"), describing their

roles in the context of a single dataset (or statistical agency). We also provide a brief overview of the rationale and methods for combining multiple data sources and describe some of the statistical approaches that have been developed to combine data and estimates from different sources. We conclude with a discussion of implications for the environment that would be needed for utilizing multiple data sources in the future, including statistical methods of combining data sources, mechanisms for research access, and approaches for protecting privacy and preserving confidentiality.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Methodological Brief: Experiments and Innovations in Exit and Election Polling
Question-Order and Third Party Candidate Support in the 2016 Presidential Elections

Eran Ben-Porath *SSRS*

Sarah Dutton *CBS News*

Melissa Herrmann *SSRS*

Jennifer De Pinto *CBS News*

The 2016 presidential election posed a challenge to polling organizations surrounding the inclusion of third party candidates in their polls. Among these challenges was the question of which candidates to include in the ‘horserace’ question, and in what order to ask these questions. Many organizations chose to ask the standard two-way horserace questions (Trump v. Clinton) as well as a three-way or four-way question (including Gary Johnson and Jill Stein). An experiment conducted over multiple waves during the months leading to the 2016 presidential elections demonstrates the impact of asking both questions, and specifically, the effect of question order. Respondents were randomly assigned to being asked the three-way horserace question before the two-way, or to being asked the two-way question first. Doing so, allowed us to test whether Johnson’s relatively poor performance in polls was, in part, a product of question order, as he and his supporters contended. The results indicate that the explicit inclusion of third-party candidates affected reported voting intent (for third party candidates as well as for the two main candidates), but the order in which this was asked produced minor and inconsistent effects. The third-party candidate did not consistently benefit or lose by the order in which these questions were asked. Nor did either of the main candidates. In our study, we detail the patterns of the effect of including third-party candidates, and focus on the consequences of question-order in this regard, overall, and for specific populations. The findings can be helpful to inform practices regarding third-party candidates in future election polls.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Methodological Brief: Experiments and Innovations in Exit and Election Polling

Testing A New Methodology for Exit Polling: A National, Panel-Based Experiment

Becky Reimer *NORC at the University of Chicago*

Becky Reimer *NORC at the University of Chicago*

Jennifer Benz *NORC at the University of Chicago*

Trevor Thompson *NORC at the University of Chicago*

Liz Kantor *NORC at the University of Chicago*

Rosalind Koff *NORC at the University of Chicago*

J. Michael Dennis *NORC at the University of Chicago*

Emily Swanson *The Associated Press*

David Pace *Associated Press*

Traditional Election Day exit polls provide important insight into voter behavior, perceptions, and demographics. However, they are logistically intensive, expensive, and have become increasingly difficult to field over time with increasing rates of early and absentee voting. In addition, growing uncertainty around the relationship between exit poll nonresponse and partisanship provides a unique challenge when weighting the results. To investigate an alternate exit poll methodology using online and telephone surveying, The Associated Press-NORC Center for Public Affairs Research conducted a national exit poll experiment for the 2016 presidential election using NORC's nationally representative AmeriSpeak® panel. The results highlight similarities and differences in the findings of the traditional exit poll and this national panel-based study, including reported voting behavior as well as voter attitudes and opinions. The panel-based data collection included an embedded experiment, with half of the sample receiving survey invitations starting the week before the election, and half receiving invitations only on Election Day. The results of this experiment address the potential benefits of extending data collection ahead of Election Day to capture absentee and early voting.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Methodological Brief: Experiments and Innovations in Exit and Election Polling

Exit Polling and Geolocation Technology: Assessing the feasibility of sending people a survey on their smartphones immediately after they vote

David Sterrett *NORC at The University of Chicago*

Jennifer Benz *NORC at the University of Chicago*

Rene Bautista-Martinez *NORC at the Univ of Chicago*

Rosalind Koff *NORC at the University of Chicago*

David Pace *Associated Press*

Emily Swanson *The Associated Press*

Trevor Thompson *NORC at the University of Chicago*

In order to explore potential improvements in exit polling, this study tested the feasibility of using geolocation technology to send voters a survey on their smartphones after they leave their polling places. The Associated Press and NORC at the University of Chicago, through funding from the Knight Foundation, conducted the feasibility study with participants in Florida during the 2016 presidential election. NORC leveraged its nationally representative AmeriSpeak panel supplemented with opt-in sample to test people's willingness to complete an exit poll on their smartphone. About 1,000 participants were invited to download the SurveySwipe app, which uses geolocation technology to send people surveys when they enter certain locations. The latitudes and longitudes of participants' early voting and Election Day polling places were programmed into the app and people received a short survey via the app after entering their polling place. The results highlight people's general willingness to participate, the download rate of the app, and the completion rates for the survey. The analysis illustrates the demographic profiles of those who cooperate and download the app. In addition, the study findings highlight various challenges and opportunities, from constructing geofences to gaining cooperation, when using geolocation technology to conduct rapid response surveys.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Methodological Brief: Experiments and Innovations in Exit and Election Polling
Polling Minor Party Candidates: An Experimental Approach

Dan Cassino *Fairleigh Dickinson Univ*
Peter Woolley *Fairleigh Dickinson University*

Minor party candidates present a unique challenge for pollsters. Failing to read the name of a minor party candidate may lead to support for the candidate being underrepresented; reading the names of minor party candidates may lead to overstated support for the candidates. These decisions become even more important given the use of survey data to determine access to debates, and, in some cases, ballot positions. Building on earlier published research on a gubernatorial election, we make use of a series of RDD likely voter studies employed during the 2016 US Presidential election. In these studies, respondents were asked about their preferences in the election in multiple ways, and in various experimental designs. The results suggested that support for the minor party candidates was far lower than estimated by conventional methods. In essence, much of the support apparently received by minor party candidates was an artifact of question design. The results suggest an alternative way to measure support for minor party candidates that better predicts actual results in an election.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Methodological Brief: Experiments and Innovations in Exit and Election Polling
Finding The Swing Voter: Definitions and Survey Methods for Voter Classification
Charlotte Ann Swasey *Civis Analytics*

This paper proposes a theory mapping emotional reactions to political information onto voter decisionmaking and then further onto measurable survey response. Using on-line processing based in emotion, voters form affective summaries about candidates, which store previous information as an emotional response. The act of voting is treated as a single realization of a probabilistic event, with the relative probabilities of each vote option being an expression of the affective summary. These summaries are expressed as warmth or feeling towards each candidate, which can be captured using the ANES Feeling Thermometer scales. A metric of the difference between the scores given to the Republican and Democratic candidates is used, based in the work of William Mayer. This metric suffers from significant survey error, but is related to party ID and expressed vote choice, as well as demographic factors and perceived efficacy. Feeling thermometer responses are found to carry meaningful information about a respondent's relationship to the election and candidate preference.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Methodological Brief: Experiments and Innovations in Exit and Election Polling
Studying Interviewer Effects in Ratings of Donald Trump

Lydia Saad *Gallup*
Rajesh Srinivasan *Gallup Inc*
Jeffrey Jones *Gallup*
Stephanie Marken *Gallup*
Frank Newport *Gallup, Inc.*

Donald Trump's presidential campaign will long be remembered for the controversial statements he made about women, Mexicans, immigrants, and Muslims. The resulting criticism Trump received from his opponents and in the media likely contributed to his low favorable ratings as measured by Gallup Daily tracking from July 2015 through the 2016 election. Leveraging the enormous sample sizes involved in Gallup Daily polling, this paper takes a detailed look at whether Trump's -- and for comparison's sake, Hillary Clinton's -- favorable rating among men, women, blacks, Hispanics and other groups differed depending on the race/ethnicity or gender of the interviewer. The hypothesis is that, due to social desirability concerns, respondents interviewed by a female, black or Hispanic interviewer may have been more reluctant to express a positive opinion of Trump than those interviewed by a male or white interviewer. The paper focuses on several discreet time periods during the campaign to address whether, to the extent there were interviewer effects, they were more pronounced at times when Trump's controversial positions were dominating the news.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Methodological Brief: Experiments and Innovations in Exit and Election Polling
Election 2016 - Order Effects and the Horserace Question

Elizabeth Steele *Gallup*

Question order effects are well-documented in the survey literature. For this reason, in political polling, many organizations place the horse race questions immediately after, or closely after, any household selection or screener problems in a survey. As part of a methods test, Gallup conducted an election survey from November 1, 2016 – November 8, 2016. The survey included a module that tested order effects on the horserace question. In this module, respondents (N = 31,453) were randomly assigned to one of five conditions. The control condition presented the traditional horserace question. A presidential approval question and a question about satisfaction with the way things are going in the United States were asked later in the survey. In the second condition, presidential approval was presented immediately before the horserace question. In the third condition, satisfaction with the United States was presented immediately before the horserace questions. Conditions four and five presented both the presidential approval and the satisfaction with the United States questions before the horserace question. In condition four, the presidential approval question was presented before the satisfaction with the United States question, and in the fifth condition, the satisfaction with the United States question was presented before the presidential approval question. The variation (or lack thereof) in the horserace responses will be analyzed between conditions. Additional analysis will look at the this potential order affect between those reporting strong support for their chosen candidate and those reporting moderate support for their chosen candidate.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Expanding Our Horizons: Attitudes and Opinions from across the Globe
National Pride across Countries and Time

Tom W. Smith *NORC at the University of Chicago*

The International Social Survey Program (ISSP) has conducted cross-national studies of national identity in 1995, 2003, and 2013. Altogether over 50 countries were covered. Included in these studies were two measures of national pride: 1) five agree-disagree items on general national pride and 2) a ten-item scale of pride in specific domains (e.g. economic achievement, military, sports). This study examines cross-national differences in level of pride, trends in level of pride within countries, and socio-demographic, sub-group differences especial by age, sex, education, and majority-minority status (e.g. White vs. Black in the US, Flemish vs. Walloons in Belgium). In almost all countries national pride is higher in older generations, among the less educated, and among members of the cultural majority. Age-period-cohort analysis indicates that inter-generational change is the main cause for the lower pride among the newer cohort. Theories of globalization and historical events are used to explain that inter-generational change.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Asking Questions About Asking Questions: Developments in Cognitive Interviewing

Running Surveys with Businesses in Japan: Comparison of Pre-Testing Methods Using Expert Reviews and Cognitive Interviews

DanDan Zhang *Twitter*

Writing survey questions and running large scale survey study in a country or domain that is less familiar to the researcher can be daunting. To ensure questions are easy to understand and relevant to the local context, survey experts have been using questionnaire pre-testing techniques to help identify any potential issues early on. There are several different evaluation methods in questionnaire pre-testing, such as: expert reviews, focus groups, cognitive interviews, and experiments; and there are several standards to assess different pre-testing methods including content standards, cognitive standards and usability standards (Groves et al. 2009). In this study, we will share the learning of using expert reviews and cognitive interviews to pretest questionnaire for a large scale survey study with small and medium businesses in the Japan market. We'll discuss how these two methods perform against the three standards. In addition, we will discuss the pros and cons of using these two different methods in an organizational setting where other factors such as cost, timeline, and stakeholder's involvement are also important considerations.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Survey Mode in Election Polling

Survey mode and social conservatism: a multi-study investigation

Patrick Sturgis *Univ. of Southampton*

Johnny Heald *ORB International*

A notable feature of recent election polling has been the tendency for polls conducted using different modes of interview to produce systematically different estimates of vote intention. For instance, in the run up to the 2015 UK General Election, the phone polls gave systematically higher readings of Conservative support than online polls, while online polls produced higher estimates of support for the UK Independence Party (UKIP). The difference between phone and online modes was even more pronounced during the EU Referendum campaign, with phone polls showing a substantial lead for Remain and online polls suggesting that the race was too close to call, with Leave holding a narrow lead based on point estimates. Systematic mode differences have been apparent in polling during the 2016 US Presidential race, with Clinton holding a bigger lead over Trump in phone than in online polls. In this paper, we investigate one putative cause of the mode effect in election polls: that online panels contain a systematically higher composition of socially conservative voters compared to phone samples (Singh and Kanagasooriam, 2016). We analyse response distributions for a set of attitudinal indicators measuring social conservatism across two phone polls, two online polls, and a face-to-face survey that were fielded in the UK shortly before and after the EU Referendum. Our results show striking differences in the levels of socially conservative attitudes across the five data sets, although not in a way that is consistent with the hypothesis that mode differences in vote intention are driven by skews in sample composition on this attitudinal dimension.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Survey Mode in Election Polling

Results of a Multi-mode Design on Pre-election Surveys

Jacqueline Redman *Floyd Institute*

Scottie Thompson *Floyd Institute*

Berwood Yost *Floyd Institute*

Pre-election polls are critical tools used by both partisan and non-partisan groups to direct, shape and understand campaign dynamics and election outcomes. Despite polling's importance, many have started questioning its accuracy, reliability, and even its future as changing data collection methodologies have been forced on the industry because of rising costs and changing communication technologies. Although traditional sampling and survey methodologies, such as Random Digit Dialing (RDD) and live-interviewer telephone interviewing, are costly and raise the potential for biases due to coverage and non-response error, they also have notable strengths. Multi-mode designs may allow methodologists to combine the strengths of traditional telephone methods with the strengths of emerging data collection tools to produce more reliable, more accurate, and more efficient pre-election polls. This paper takes an in-depth look at the effects of adopting a multi-mode interviewing design (web and telephone) on survey efficiency, response rates, sample representativeness, and survey estimates in pre-election surveys by evaluating the results of multi-mode surveys conducted with registered voters in Pennsylvania during July, August, September, October, and November, 2016.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Expanding Our Horizons: Attitudes and Opinions from across the Globe
Afghan Futures: Updates from the Latest ACSOR/D3 Polls in Afghanistan
Matthew Warshaw *D3 Systems, Inc.*

Through the presentation of data from our on-going “Afghan Futures” polling series, we propose to present new data on Afghan attitudes towards the power sharing government, the reduction in international forces and international aid spending, a resurgent Taliban, and the continued presence of ISIS sympathizers in Afghanistan. We will look at the results from our recent national surveys and discuss what comes next for Afghanistan. ACSOR Surveys and D3 Systems will present survey data from our most national polls in Afghanistan, to examine the above mentioned issues as well as other topics such as migration and the economy in Afghanistan. The goal of our “Afghan Futures” project is to continue to enrich public understanding of Afghanistan, as well as to share insights into the methodological and operational challenges of conflict-zone research. The project began in 2010 and includes results from eight national face-to-face surveys. We will have new data from the 9th wave of the project to present at AAPOR. Interviews are conducted in Dari and Pashto among a random national sample of an average of 2,000 respondents in each wave covering all thirty-four provinces of Afghanistan.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Asking Questions About Asking Questions: Developments in Cognitive Interviewing

Screening for Specific Experiences: Striking the Balance between False Positives and False Negatives

Mandi Martinez *U.S. Census Bureau*

Mary Davis *U. S. Census Bureau*

Amber Henderson *U. S. Census Bureau*

For many surveys that aim to gather in-depth data on particular experiences such as crime victimization, one of the first challenges is to identify the right set of individuals for a detailed battery of follow-up questions. Various strategies may be employed to identify the target set of respondents for the follow-up questions. Some surveys cast a wide net using screener questions that include a broad range of experiences that could qualify as the target experiences without using specific labels. For example, questions aimed at measuring prevalence of domestic violence include prompts designed to capture experiences that some respondents may not, themselves, consider to be a crime. The aim is to provide stimuli on a wide range of experiences in order to define the scope of experiences that could “count,” and to jog respondents’ memories of such experiences. Further questions are asked to determine whether the reported incidents meet all of the criteria required for the experiences being measured. Other studies use more direct language that is likely to align more closely to the experience under study. The challenge is to identify the most appropriate words, terms and strategies in the screener questions to minimize both false positives (where someone is included in the universe who should not be) and false negatives (where someone who belongs in the universe is left out). In this presentation, we describe how cognitive interviews allow researchers to develop a set of screener questions for the survey goals that aims to strike the right balance to minimize both false positives and false negatives. This is done by asking follow-up probes on respondent experiences that allow the researcher to identify issues affecting potential measurement error in the survey. We illustrate this through the testing of two survey supplements to the National Crime Victimization Survey.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Using Advance Notification to Effectively Increase Participation

Increasing census self-completion rates for hard-to-enumerate populations: A social marketing journey

John Beler *Statistics Canada*

Daniel Houle *Statistics Canada*

The communications campaign for the 2016 Census in Canada was both ground-breaking and extremely successful. Nearly 9 out of 10 Canadian households completed and returned their questionnaire without any assistance from Statistics Canada staff. A self-response rate of 88.8% would not have been achievable without a considerable increase in self-completion by harder-to-enumerate audiences. The Integrated Communications Strategy (ICS) for the 2016 Census of Population was designed as a social marketing campaign. The emphasis was placed on research, segmentation, targeting, and positioning. The Census Communications team adopted the main tenet of the Elaboration Likelihood Model (ELM), which outlines the need to leverage both direct and peripheral communication channels to draw an audience's attention to targeted messaging. The overall targeting strategy consisted of full market coverage, and the use of differentiated marketing to reach audiences that had proven more difficult to enumerate. Statistical analyses guided the design, implementation and monitoring of census communications activities. The Integrated Communications Strategy employed behavioural and sociodemographic segmentation to identify hard-to-enumerate audiences. The goal of the segmentation was to understand which areas are more or less predisposed to self-respond, and what key sociodemographic characteristics had to be considered in the execution of communication activities. The overarching objective for all marketing activities was also to maximize engagement. Social media content, in particular, was designed to be topical, timely and targeted. Outreach activities were proactive, systemized and targeted. A Community Support Toolkit was developed and distributed to over 4,000 community-based organizations (CBOs), associations and municipalities across Canada. 600 organizations and municipalities readily made use of social media calendars placed at their disposal. This session will offer an outline of the ICS, including lessons learned and best practices identified in the course of 2016 Census of Population.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

The MISSING Link: New Methods for Survey Data Imputation

Visibility Imputation for Respondent-Driven Sampling

Katherine R McLaughlin *Oregon State University*

Mark SHandcock *University of California, Los Angeles*

Respondent-driven sampling (RDS) is a network sampling method designed to access hard-to-reach populations who cannot be sampled by traditional methods. A common application is to gather information about populations at high-risk for HIV/AIDS throughout the world to aid in resource allocation for treatment and prevention. In RDS, study participants recruit a capped number of their peers to enroll, making the sampling mechanism unknown to researchers. Design-based inference is made possible by statistically modeling recruitment over an underlying social network in the population. Calculation of inclusion probabilities depends on the self-reported personal network sizes (degrees) of participants, which may be misreported or biased for a variety of reasons. Intentional misreporting due to the stigmatized nature of many sampled populations, heaping, and lack of understanding of the questions can lead to poorly specified degrees. In this talk, I discuss RDS implementation, several ways the personal network size variable is attained, and introduce a measurement error model to impute individual visibility based on each participant's self-reported network size, number of recruits, and time to recruit. These imputed visibilities can also be thought of as a way to smooth the degree distribution and bring in outliers, as well as a mechanism to deal with missing and invalid network sizes. They can be used in place of degree in existing RDS estimators. Finally, I demonstrate the performance of inference for the visibility distribution on a population of men who have sex with men (MSM) from Prishtina, Kosovo in 2014.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

The M_SSING Link: New Methods for Survey Data Imputation

Administrative Records Use for Item Imputation

Andrew D. Keller *U.S. Census Bureau*

Item imputation has been a much researched topic in survey methodology. In general, imputation models take advantage of reported respondent data to make inferences about missing data. To aid in item imputation, administrative records (AR) linked to survey data can supplement existing data collected from respondents. The Census Bureau has been investigating applications of AR during the decennial census. As part of that research, we have explored possible uses of AR for item imputation of missing person and housing unit characteristics on census responses. This paper proposes methodologies that incorporate information from AR for item imputation of census characteristics. Using the 2010 Census results, it then contrasts imputed characteristics against results implementing those proposed methodologies integrating AR. The goal is to understand how an imputation scheme combining AR might affect measurement of subpopulations.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Using Advance Notification to Effectively Increase Participation

To Notify or Not To Notify – The Impact of Prenotification Letters in Establishment Surveys on Response Rates and Data Quality

Thomas Brassell *ICF International*

Brian Orleans *ICF International*

Robert Tortora *ICF International*

James Dayton *ICF International*

Andrew Blevins *Pew Charitable Trusts*

Theron Guzoto *Pew Charitable Trusts*

Alison Shelton *Pew Charitable Trusts*

John Scott *Pew Charitable Trusts*

Sarah Spell *Pew Charitable Trusts*

Previous research has found prenotification letters improve response rates for household surveys regardless of data collection methodology (i.e., web, mail, CATI) (Dillman, Smyth and Christian, 2014). Last year, we explored their efficacy in establishment surveys by mailing a portion of our small business sample prenotification letters. We reported that after seven weeks of dialing, we found significantly higher cooperation rates and significantly lower refusal rates among those respondents that had been mailed a prenotification letter. However, interestingly, the vast majority of respondents that had been mailed a prenotification letter did not recall receiving the prenotification letter. The current study follows-up on our previous findings by expanding the scope of the research to include a third condition where respondents were informed that we had mailed a letter when in fact no letter had been sent. Response rates, cooperation rates, and refusal rates will be compared across all three conditions. We hypothesize that respondents that received a prenotification letter will have significantly better participation metrics than the other two conditions. In addition to assessing the impact of prenotification letters on respondent participation, we will assess its impact on data quality. Additional information is generally conveyed in prenotification letters regarding the study than may be discussed over the phone. Furthermore, if the receipt of the letter predates completion of the survey by several days, the information provided in the letters may influence responses as thoughts on the topic have had time to mature. We will assess for differences in data quality (e.g., refusal and don't know responses), as well as attitudinal differences on the topic of small business retirement plans. We hypothesize that there will be no significant attitudinal differences, but that data quality will be superior among those respondents that received a prenotification letter.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Survey Mode in Election Polling

Learning from the 2016 General Election Presidential Debates: What difference does mode make?

Kenneth Winneg *Univ. of Pennsylvania*

Kathleen Hall Jamieson *Univ. of Pennsylvania*

Eran Ben-Porath *SSRS*

Over four decades of research show that presidential debates provide an opportunity for voters to learn about the issues and form opinions about the character of the candidates running for office. Knowledge of candidate issue positions plays a role in vote choice. Does mode of interviewing and pre-recruitment to view a debate effect knowledge measures? This paper will compare differences in candidate knowledge among viewers of the first presidential debate on September 26, 2016 between Democrat Hillary Clinton and Republican Donald Trump, drawn from two modes of data collection: 1) probability based online panel of more than 5,000 respondents, and 2) a cross-section of 1,004 U.S. adults (national RDD telephone sample) including 727 viewers of the first debate. Both post-debate surveys were conducted beginning the day following the first debate in order to measure the impact of post-debate news coverage immediately following the debate. The online study was in the field for three days while the telephone study was in the field for six days. For the online sample, we recruited a probability based panel of U.S. adults to watch the first debate. In the online sample, a representative subsample (n=2,520) was asked about positions of the candidates before and after the debate, along with candidate favorability and vote intention. Another representative subsample (n=2,625) was only asked the issue knowledge questions post-debate. The telephone sample, was conducted post-debate only and therefore respondents were not exposed to the issues prior to the debate. Preliminary results show evidence that debate viewers were knowledgeable of key issues discussed in the debate, regardless of mode, and the self-identified debate viewers in the telephone sample were significantly more knowledgeable than self-identified non-viewers in the phone sample. This research will contribute to our understanding of mode effects and priming in the context of presidential debates.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Asking Questions About Asking Questions: Developments in Cognitive Interviewing

Apples and Oranges: What is the right question when comparing Web Probing and Cognitive Interviewing?

Jennifer Edgar *Bureau of Labor Statistics*

Paul Scanlon *Centers for Disease Control and Prevention*

Web probes have been increasing in popularity over the past several years, being used in a variety of ways to evaluate questions (Braun 2008; Behr, Braun, Kazczmirek, and Bandilla 2012; Edgar, 2012 and 2013; Fowler et al, 2015; Murphy, Edgar, and Keating 2014). To understand this new methodology, several studies have compared the information collected with web probes to information collected with traditional, laboratory cognitive interviews (Edgar, Murphy, and Keating 2016; Holzbert, Clark, Moralez and Childs, 2016; Meitinger and Behr, 2016; Neuert and Lenzer, 2016). In this presentation, we argue that how web probing compares to traditional cognitive interviewing is the wrong question. Instead, we propose that researchers consider web probing a related, but unique, evaluation tool with its own strengths and weaknesses. Simply replicating the traditional cognitive interviewing approach online does not leverage the strengths of web probing, and may limit the potential benefits of the approach, particularly if the potentially large, and geographically diverse, sample size is limited by the amount of time required to analyze the responses. Reviewing the types and forms of web probes that have been used (e.g. expansive, categorization, open-ended and closed-ended; Behr 2016; Fowler 2016; Scanlon 2016), we will discuss the relative strengths and weaknesses of each, with the goal of starting an inventory of probe types to inform their research design decisions. Providing examples of the probe types, we aim to place this methodology in the context of the overall question evaluation process, and to inform the ongoing conversation about the role of web probing in question evaluation and how it relates, or not, to traditional cognitive interviewing. The presentation will also identify the issues that have yet to be addressed with the technique, and the other opportunities for further research.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Using Advance Notification to Effectively Increase Participation

Reducing Nonresponse: A Randomized Experiment on Advance Letters in Two National Multi-Mode Establishment Surveys

Lauren Harris-Kojetin *National Center for Health Statistics*

Celia Eicheldinger *RTI International*

Manisha Sengupta *National Center for Health Statistics*

Melissa Hobbs *RTI International*

Angela Greene *RTI International*

Although research finds advance notification letters have a positive effect on response rates (Chiu & Brennan, 1990; de Leeuw et al., 2007; Dillman 2000; Edwards et al., 2009; Fox et al. 1988; Heberlein and Baumgartner 1978, 1981; McColl et al., 2001), some studies find no significant difference (Australian Bureau of Statistics, 2002; Cycyota & Harrison, 2002; Drummond et al., 2008; Hammink et al., 2010; Shiono & Klebanoff, 1991). Most research on advance letters has been with population surveys, with less on establishment surveys (e.g., Cycyota and Harrison, 2002). Past research examined advance notification postcards versus letters (Hembroff et al., 2005; Richardson, 2011), but not the message itself. This experiment examined advance letters in a multi-mode establishment survey, specifically whether a simpler advance letter message—or no letter—helps to increase participation, especially earlier participation using less expensive modes. The 2016 National Study of Long-Term Care Providers, a biennial study sponsored by the National Center for Health Statistics (NCHS), included two establishment surveys, one of adult day services center directors and another of assisting living community directors. In prior years, NCHS included in its advance letter technical references to laws that empower NCHS to conduct the survey, assuming that respondents will be more willing to participate if they can see NCHS's legal authority. However, the legal references may cause respondents to lose interest and avoid the subsequent questionnaire mailings. Cases were randomly assigned to one of three groups (conventional technical advance letter, simpler advance letter, or no advance letter-control group). The protocol included three questionnaire mailings, a thank you/reminder letter after the second questionnaire, then computer-assisted telephone interviewing for non-respondents. Preliminary results suggest that for both sectors, the simpler letter produced higher participation than the technical letter which produced higher participation than no letter. Final logistic regression results will be reported.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Expanding Our Horizons: Attitudes and Opinions from across the Globe
Media Use in the Middle East, 2013-2017: Findings and Reflections from a Longitudinal Study of Media Behaviors and Political Attitudes in Six Arab countries

David Krane *Nielsen Consumer Insights*
Justin Martin *Northwestern University in Qatar*
Everette Dennis *Northwestern University in Qatar*
Robb Wood *Northwestern University in Qatar*
Mariam Saeed *Northwestern University in Qatar*

Northwestern University in Qatar has surveyed media use and political and cultural attitudes in the Middle East every year since 2013. In a few years, some significant changes in media habits and public opinion have been tracked. Internet penetration has soared in Egypt and Lebanon, Twitter use has fallen sharply across the region, Facebook use has plummeted in Saudi Arabia and Qatar, and newer services such as Snapchat and Instagram are ascendant in many countries. Findings on the languages adults use to communicate and access mass media suggest that reliance on and use of Arabic media is rising in some countries, while English use is falling. Among some of the public opinion trends in the data, fewer nationals express confidence in the credibility of news media in their respective countries. More internet users are concerned about online surveillance by companies than by governments. Fewer adults say people should be able to criticize governments on the internet. The study, *Media Use in the Middle East*, surveys nationally representative samples of adults each year in Egypt, Lebanon, Qatar, Saudi Arabia, Tunisia and the U.A.E., and includes data from more than 34,000 respondents since 2013. Data are collected in face-to-face interviews in all countries except Qatar, where random-digit dialing is used. The questionnaire is administered in Arabic, English, and French. This presentation will highlight trends from Northwestern's media use surveys in three areas: social media use, language use, and attitudes about freedom of expression and censorship. The contribution will share findings about what tools and platforms people in the region are using to communicate and consume information, changes in languages used to do so, and complex, seemingly contradictory attitudes about Western media.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Using Advance Notification to Effectively Increase Participation

Impact of Email Tone on Response Rates among Young Lesbian, Gay, Bisexual and Transgender (LGBT) Respondents

Patricia LeBaron *RTI International*

Kristine Wiant *RTI International*

Gretchen McHenry *RTI International*

Leah Fiacco *RTI International*

Tesfa Alexander *US Food and Drug Administration*

Is a formal tone always the most effective for soliciting participation in a web survey? The Food and Drug Administration's Research and Evaluation Survey for the Public Education Campaign on Tobacco among LGBT (RESPECT) project built a panel of 18 to 24 year old respondents through both in-person and social media screening and recruitment. Surveys are conducted twice a year for two years. Participants received an email inviting them to participate in a longitudinal web survey as well as reminder emails throughout the data collection period. Given both the young population and the informal settings in which they were recruited (gay and lesbian bars), we hypothesized that the population might be more likely to participate in the survey when the email invitations and reminders were informal in tone. We constructed two sets of invitations, one using typical formal language and tone, and the other using informal language and tone. For example, the subject line of the formal email was "Your invitation to the RESPECT Study." The informal email's subject was "Top 6 Reasons to respond to the RESPECT survey." Respondents were randomly assigned to receive either all formal messages, including invitation and reminders, or all informal messages. This presentation will discuss the relationship between response outcomes and email tone. Relationships between invitation type and other measures of participation, such as breakoffs, will also be explored. This research will contribute to knowledge about effective ways to reach out to young survey respondents, especially those in the LGBT community.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

The M_SSING Link: New Methods for Survey Data Imputation

Non-Parametric Multiple Hot Deck Imputation for Surveys

Natalie Jackson *Huffington Post* / *POLLSTER.COM*

Jeff Gill *Washington University*

Skyler Cranmer *Ohio State University*

Nonresponse is a common problem in surveys on items of interest as well as demographic items used for weighting and analysis, but it is often not addressed correctly. In many cases the relevant variables are discrete rather than continuous, which limits the use of most multiple imputation techniques that impute continuous values. Researchers who use these techniques are forced to make rounding decisions that may be inappropriate for the data, such as determining how to round an imputed value of 0.49 or 0.51 for a variable that was measured as 0 or 1. Multiple hot deck imputation techniques (Cranmer and Gill 2013) offer a solution by imputing discrete values where the missingness is discrete, and as such resolves the problems with using other imputation techniques on classic survey data. This paper applies shows how this method can be applied to demographics of a sample match Census data and demonstrates how close pre-election poll estimates can be to electoral outcomes when missing data is dealt with correctly.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

The M_SSING Link: New Methods for Survey Data Imputation

Indirect Estimation of Race/Ethnicity for Survey Respondents who do not report Race/Ethnicity

Marc Elliott *RAND Corporation*

Amelia Haviland *Carnegie Mellon University*

Katrin Hambarsoomian *RAND Corporation*

Jacob Dembosky *RAND Corporation*

Samuel C Haffer *Centers for Medicare & Medicaid Services*

Even when self-reported R/E is a key analytic variable requested on a survey, a non-random subset of survey respondents may omit this item. In the Medicare CAHPS survey of patient experiences, we use self-reported race/ethnicity, which is present in most but not all cases. Here we imputed race/ethnicity using the validated Medicare Bayesian Improved Surname and Geocoding (MBISG) method for indirectly estimating race/ethnicity for survey respondents who did not report race/ethnicity. We sought to understand how often self-reported race/ethnicity is missing in CAHPS data, how often indirect estimates of race/ethnicity are available when self-reported race/ethnicity is missing, and whether different racial/ethnic subgroups are more or less likely to report their race/ethnicity. Self-reported race/ethnicity was missing in approximately 4.5% of all cases (26,987 of 616,777 respondents). This rate of missingness for race/ethnicity does not appear to be a function of the content of the item, but to reflect general factors regarding item nonresponse, including breaking off from telephone interviews. Indirect estimates were available for 95.6% of the cases in which self-reported race/ethnicity was not available. We estimate that 7.7% of Blacks, 5.9% of Hispanics, 5.2% of Asians/Pacific Islanders (API), and 3.8% of non-Hispanic Whites did not self-report their race/ethnicity ($p < 0.05$). Indirect estimation of race/ethnicity for those respondents who did not self-report race/ethnicity had a minimal effect on the overall estimated distribution of respondents by race/ethnicity. It also results in a 4% increase in the overall sample size (9% for Blacks and 7% for Hispanics) and corresponding small gains in statistical power. There may be some benefit to using our BISG method to indirectly estimate race/ethnicity in those cases where survey respondents do not self-report race/ethnicity, especially if it is missing at high and varying rates.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Asking Questions About Asking Questions: Developments in Cognitive Interviewing

Is a Proxy Response Good Enough? Using Paired Cognitive Interviews to Assess the Accuracy of Proxy Responses

Amber J. Henderson *U.S. Census Bureau*

Mary C. Davis *U.S. Census Bureau*

Jenna Fulton *U.S. Census Bureau*

Conducting a survey interview with a single household member who can accurately provide proxy reports for all other household members can offer cost benefits relative to conducting a survey interview with each member of the household individually (Blair, Menon, and Bickart, 1991). This paper evaluates the feasibility of proxy interviewing through the use of paired cognitive interviews conducted with related or unrelated members of the same household, using an approach described by Blair et al. (1991). This paper uses two recent cognitive interview projects from surveys on social attitudes and behaviors, and computer and internet use as case studies from which we provide guidelines for conducting paired cognitive interviews to evaluate the feasibility of using proxy response, including recruiting and scheduling procedures for paired interviews with related and unrelated respondents, and potential challenges. This paper will also examine differences between self and proxy response for questions subject to social desirability bias, such as those on voting and donating to charities or political organizations, as well as internet and technology use. In addition, we will present suggestions for analyzing data from paired cognitive interviews in order to determine the feasibility of using proxy interviews. This includes quantitative analysis, such as by calculating the rate of matching or near-matching responses, and analysis of qualitative feedback, such as identifying characteristics of questions that are easier or more difficult for proxies to answer, and identifying characteristics of household members or relationships that facilitate or impede the ability to accurately respond by proxy. Finally, this paper will discuss overall benefits and shortcomings of using paired cognitive interviews to evaluate the usefulness of proxy interviewing, and potential applications for this research design.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Survey Mode in Election Polling

Fifty, Nifty United States

Joseph Zappa *Ipsos Public Affairs*
Kaitlyn AMcAuliffe *Ipsos Public Affairs*
Clifford Young *Ipsos Public Affairs*

National tracking polls and state-specific polls have long been a popular method of measuring national voting intention for elections in the United States. Unfortunately, traditional methods of polling such as random digit dialing have increasingly suffered from growing non-response rates. Consequently, pollsters have capitalized on technological advancements to modernize their practices: Collecting large non-probability based samples via the internet has become highly cost-effective, and increasing computational power has made it possible to use Bayesian methods that increase the accuracy of forecasts by marrying national and state level data. However, the unexpected outcome of the 2016 election has forced pollsters to examine what caused their predictions to err. We hypothesize that one plausible explanation is that pollsters overlook the importance of correlations among state polls. State polls are traditionally fielded in battleground states, but not in clusters of states. As such, state polls fail to measure what is often a strong correlation among states, which was of central relevance in accurately measuring rural America's voting intention in the 2016 election. In 2016, Ipsos launched a high volume, national mega-poll that measures demographic characteristics such as party identification, the respondent's likelihood to vote, and Presidential voting intention. Using weight efficiencies as a proxy for population representivity and cost per interview, we examine the relative accuracy and efficiency of three different methods (national tracking poll, state level polls, and the mega-poll) for measuring the voting intention of a state geography. We conclude that with careful survey and weighting design, non-representative mega-polls in conjunction with prior voting history offer an affordable alternative to traditional polling in elections in the United States.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Asking Questions About Asking Questions: Developments in Cognitive Interviewing

Conducting Joint Cognitive and Usability Testing of Messaging about Privacy and Confidentiality Protections

Jessica L. Holzberg *U.S. Census Bureau*

Upon receiving a request to participate in a survey, some sample members may turn to the Internet for further information before making the decision to respond. Thus, it is important to ensure that messaging about the survey available online is both easy to understand and easy to access. This may include information such as the survey's purpose, length, and privacy and confidentiality protections. Traditionally, researchers at the U.S. Census Bureau would conduct in-person cognitive and usability testing separately to assess the understanding and accessibility of this messaging, respectively. Performing cognitive and usability testing independently is more time-consuming and expensive, and eliminates the opportunity to determine whether respondents' ability to find and understand survey messaging interact. Conducting testing in-person also introduces stricter limits on the number of respondents with whom one can test and the geographic diversity of those respondents. We conducted an online joint cognitive and usability test of messaging about privacy and confidentiality protections. Respondents were given privacy and confidentiality scenarios and then asked to locate relevant information on the census.gov homepage. We collected information on whether the respondent successfully navigated to the messaging of interest on the website and what they clicked on to get there. Following each navigation task, respondents were presented with the same privacy and confidentiality messaging and asked fixed open-ended cognitive probes. These probes were administered to all respondents regardless of whether they successfully completed the prior navigation task. This paper will provide an overview of this methodology as well as substantive findings on effective messaging about respondent privacy and confidentiality.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Expanding Our Horizons: Attitudes and Opinions from across the Globe

Receiving and Sending Remittances: Estimating the Percent of Adults Receiving/Providing Financial Support

Anita Pugliese *Gallup*

Julie Ray *Gallup*

Neli Esipova *Gallup*

Gallup World Poll conducts annual surveys in 125 to 150 countries. This research platform is an important method of tracking worldwide issues regarding migration and remittances. Gallup World Poll has collected data across more than 150 countries on 1) the percent of adults receiving financial help (in the form of money or goods) from someone within the same country, or another country, and 2) percent of adults sending financial help to someone within the same country, or another country. These data focus on percent of adults receiving/sending support (and their characteristics), which complements officially recorded flows of money. The data also shed additional light on remittances in countries -- particularly those in sub-Saharan Africa -- where little or no official data exist. We will present new data collected in 2016 on both receiving and sending financial support. Gallup's results to date in 2016 show that in over 30 countries, over 10% of the population receive support in the form of money or goods from an individual in another country. We will show specific country examples where financial support from others has increased during difficult economic times. (e.g. Yemen). Receiving financial support from others within the country is not limited to developing countries; a certain percent in some highly developed countries also receive this support.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Expanding Our Horizons: Attitudes and Opinions from across the Globe

Accentuating the positive: Australian attitudes towards United States' foreign policy, 1987-2016

Steven McEachern *Australian Data Archive*

Accentuating the positive: Australian attitudes towards United States' foreign policy, 1987-2016 A strong relationship with the United States has long been a cornerstone of Australian foreign policy, dating back to the Second World War. This relationship has endured through left and right wing governments in both countries, the Cold War, Vietnam War and Gulf Wars, and continues to this day in the form of the ANZUS (Australia, Zealand and United States) treaty. A key element of this relationship is the strong public support for the US among the Australian public. Australians have recognised the need for Australian and US collaboration on foreign policy, defence, and related matters, and been willing to support US foreign policy in most circumstances, in return for the important military and economic benefits that a strong US-Australian relationship provides, particularly in the Asian region. In this context, the election of Donald Trump, under a policy of reduced participation in bilateral and multilateral agreements in foreign policy (both military and economic) has potential consequences for Australian foreign relations with both the US and Asian countries. This paper therefore seeks to explore the current state of Australian attitudes towards the United States and US foreign policy, in the context of Australia's changing relations in the Asian region. We use three Australian national time series - the Australian Election Study (1987-2016), the ANU Poll (2008-2016) and the Lowy Poll: Australia in the World (2005-2015) - to understand the current and past attitudes of Australians, key determinants of positive attitudes toward the US, and implications for changes in attitudes under a Trump presidency.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Survey Mode in Election Polling

Do Panel Surveys Produce Good Estimates of Political Participation?

Bradley Spahn *Stanford University*

A panel survey might be expected to produce less accurate estimates of voting behavior than a cross sectional survey because panel attrition and conditioning can introduce special biases unique to panels. Remarkably, for the Pew Research Center's American Trends Panel, this does not appear to be the case. Using validated vote data on 5,503 Panelists, I find that panel attrition among non-voters is small and ignorable once weights are applied. To identify causal effects, I created a control group of panelists recruited after the 2014 general election and a treatment group of panelists that had been in the panel for nearly 2 years before the 2014 election. Using prognosis score matching, the two groups were matched on basic demographics and turnout variables from the voter file. Participating in the panel made panelists about 2 percentage points more likely to vote in the 2014 election, but this higher level of participation is counteracted by decreased over-reporting of voting among panelists. Once all these panel effects are taken into account, estimates of 2014 voting rates from long-term ATP Panelists are about the same as they would be from a simple cross-sectional survey.

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

The M_SSING Link: New Methods for Survey Data Imputation

Testing a Machine Learning Approach to Missing Data Imputation

Sarah Kelley *University of California - Berkeley*

Emerging machine learning techniques offer important new opportunities for public opinion research and in particular better utilization of rich data sources, since these techniques can successfully fit a model on thousands of variables. This offers key advantages for missing value imputation – as missing values on variable of interest can be imputed with greater accuracy using all of the respondents' other answers (rather than just a pre-selected subset, as is currently standard). Non-linear machine learning models also offer the potential to accurately input missing data when relationships with other variables are not reducible to linearity. Evidence for the efficacy of this technique imputing education and income data in the World Inequality Study data (32051 cases) is presented along with comparisons of potential models (regression with regularization, Naïve Bayes, Decision Trees, and K-Nearest Neighbors).

May 18th, 2017
4:30 PM - 6:00 PM
Concurrent Session A

Using Advance Notification to Effectively Increase Participation

What Participants Want: Motivations to Complete Surveys Among Different Sample Sources

Larry Osborn *GfK*

Jordon Peugh *SSRS*

Nicole R. Buttermore *GfK*

Frances M. Barlas *GfK*

Randall K. Thomas *GfK*

Although more surveys are being conducted than ever before, individual surveys are facing declining response rates that threaten the validity of results. The decline in participation rates may in part be due to survey researchers' failure to address the motivational needs and interests of participants. Our goal was to identify the factors that motivate respondents to complete an interview. We investigated survey design features as well as survey topics using different sample types and modes including a web-based non-probability sample of 57,104 respondents (ARF's Foundations of Quality 2 Survey) and a web-based probability sample of 2,469 respondents (GfK's KnowledgePanel). We asked participants how important they considered a set of factors that typically influence participation in surveys. For both types of online samples, we found that offering cash or reward points was most important in motivating participation, followed by an interesting survey topic, and the opportunity to express opinions. Among the least important reasons for participating were sweepstakes and being made to feel part of a community. The highest rated types of surveys for both probability and non-probability online samples were those that ask about many different topics (such as an omnibus), whereas the lowest-rated surveys were those on automotive, financial, and political topics. We will discuss these findings, comparing results by mode (online and phone) and sample type (probability and non-probability). Discussion of results will focus on how we, as an industry, might improve overall study participation and how we might tailor our communication with different types of respondents to optimize their participation.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Present and the Future of Survey Research with Voter File Data (Panel)

Present and the Future of Survey Research with Voter File Data

Masahiko Aida *Civis Analytics*

Masahiko Aida

Jonathan Robinson *Catalist*

Kristen Anderson *Echelon Insights*

Ruth Igielnik *Pew Research Center*

Scott Keeter *Pew Research Center*

Courtney Kennedy *Pew Research Center*

Nate Cohn *NYT*

Patrick Ruffini *echeloninsights*

This panel will address different potential mechanisms that led to errors of 2016 presidential election polls. The central theme of the discussion is using voter list to dissect the total errors of 2016 polls. This panel assembles experts across the Polling and Media industries to approach the problem in a holistic manner: Democrats and Republican partisan pollsters, a journalist, and non-partisan survey researcher. Political survey research is at its crossroad today. The probability sampling method emerged as a dominant paradigm of research methods after the infamous polling failure of 1936. It assigns an explicit chance of selection to each element, and on expectation, it provides unbiased estimates of population parameters by design, hence the name design-based survey. As many have noted, survey non-response correlated with outcome of interest potentially produce greater magnitude of errors when the response rate becomes lower. Moreover, the increase in cell phone only households, and stringent interpretation of TCPA are collectively killing political polling based on RDD method. The memory of polling failure of 2016 that predicted the victory of HRC over Donald Trump is still vivid, while poll after poll suggested the HRC's advantage, Trump won the election by obtaining over 290 electoral votes. There are many speculations on what went wrong, in reality, it is not easy to diagnose and run post-mortem of polling errors as many polls conducted by RDD method lack keys to cross reference against actual turnout data. Several potential factors that led to the polling failure of 2016 will be discussed in the current panel. Some papers examine the systematic patterns of survey non-response using auxiliary data. Another paper discusses misspecification of turnout models and how it impacted the vote share estimates. The final paper evaluates evidence of social desirability biases to see if polls underestimated Trump support.

Dissecting Polling Errors using Voter List with Total Error Framework

Masahiko Aida

Election day is a scary day for public opinion researchers who conduct political polling. On election day you will find out if you were wrong or correct. Conversely, because election day serves as the ultimate validation, survey researchers can investigate which parts of survey operation machinery broke. The 2016 election is an opportunity to understand the mechanism of polling errors in much deeper fashion; the question is how do we go about it? This paper uses extensive micro data from 2016 election to understand the potential causes of polling errors using political survey data collected on behalf of independent expenditure clients and coordinated communication clients. During 2016, we ran over 400 studies, completed over 544k interviews, and retained approximately 72 million call dispositions. These

survey interviews were collected using voter lists as a sampling frame, and they provide extensive information on both respondents and non-respondents, thus providing deeper insights for post-election analysis. I will organize study as three step process. The first hypothesis focuses on the unreachable sample. If there are systematic differences in attitude between respondents and non-respondents, it would have manifested as non-contact biases. Many pointed out that decreasing survey response rate (Zukin, 2015; Gelman, 2015) will increase this risk factor. The second hypothesis addresses refusals; it is plausible that Trump supporters who are skeptical of institutions refused to participate in opinion surveys, as Donald Trump openly criticized biased media and its rigged opinion polling. The final hypothesis is social desirability biases. Noelle-Neumann's spiral of silence theory (1993) suggests, minority and female respondents felt hesitant to express unpopular opinions. If true, people may respond differently to live telephone survey and self-administered web survey. This paper compares parallel web and telephone surveys from voter list frame to answer this question.

Beyond RDD, Voter List Based Polling by New York Times Upshot and Siena College

Nate Cohn

Political organizations have relied on voter registration files for political polling for a long time, but most major media organizations continue to use random-digit-dialing for political surveys. The public pollsters that do use listed samples have mainly applied traditional, R.D.D. polling techniques like--random sampling, rake weighting and self-reported turnout intention--to a listed frame. Many campaigns use other methods to draw their sample or estimate likely voter scores. In Fall 2016, the New York Times Upshot and Siena College conducted five sample registration-based surveys in Pennsylvania, Florida and North Carolina. The surveys were commissioned for journalistic reasons: the rich voter file data on respondents will strengthen our journalism on the electorate for years to come. But the surveys drew on ideas from campaign polling and political science. The survey used coverage and response rate adjusted samples to reduce the need for weighting. It blended self-reported turnout and a modeled turnout score, based on vote history. And although the result was estimated using raking weighting, a number of alternative estimation methods, including multi-level regression and post-stratification, were employed alongside every survey. Along the way, we considered a variety of practical challenges, like cell phone use, date of registration, voter file updates and educational attainment. The presentation will explain and assess the Upshot/Siena effort. It will leverage the final results and updated vote history to gauge whether turnout models or self-reported vote intention accurately represented the likely electorate. It will consider whether any of the alternative weighting or estimation schemes would have consistently offered a more accurate result. It will explain the practical challenges of pursuing this type of poll for the first time—in terms of both data availability and data management— and lessons for future public listed surveys.

Why are American Presidential Election Campaign Polls still so Variable when Votes are still so Predictable? Voter Files Can Tell Us Why.

Jonathan Robinson

In 1993, Gelman and King published their famous paper "Why are American Presidential Election Campaign Polls so Variable when Votes are so Predictable?". The paper posited a theory of 'enlightened preferences', where voters react to the campaign through the lens of the news media, constantly updating their views with the latest information. A model consistent with that theory fit historical data well, and thus the ideas was born and taught to thousands of students of political science and public opinion. But 20 plus years later, partisanship predicts vote choice among survey respondents better than it ever has, and at the same time, response rates to surveys have fallen and polls still vary. The

likely culprit here is partisan-non-response, where a respondent's likelihood to respond to a survey is differential based on partisan leanings, beyond demographics. Voter file based polling and registration based sampling, especially pulled from databases used by campaigns, contain lots of auxiliary data and powerful covariates for predicting likelihood of voting as well as vote choice. Voter files also provide a sampling frame with which to evaluate response bias of all kinds. We seek to examine partisan non-response bias in depth in this paper by working with a dataset of tracking surveys from the 2012, 2014, and 2016 election cycles to examine how political events, election cycles and geographies impact partisan non-response bias as well as developing methods for detecting and ultimately correcting them. We will also explore how partisan non-response bias impacts samples and results by relying on the rich covariate profiles of campaign databases that contain propensities to support candidates & identify with parties, as well as demographic data (in some cases modeled) on race, age, gender, marital status, and education, which will allow us to qualitatively describe how this ultimately impacts results.

The Role of Commercial Voter Files in the Study of Elections

Ruth Igielnik

The recent rise in the use of “voter files”-- digital databases built on the official government voting records of individuals -- by political pollsters and campaigns has sparked increased innovation in how campaigns use and access data. These commercially-available databases are usually matched with data from consumer records, political organizations, the U.S. Census and other sources and are marketed as providing a rich and comprehensive record for nearly every American adult. This study will describe the current landscape of commercial voter files and examine how they are being used, how well they cover the adult population and the types and accuracy of information available in them. Voter files available through commercial vendors often fall into three categories: those that work primarily with Democrats and progressive groups, those that work primarily with Republicans and conservative groups, and non-partisan vendors. To assess the range of available data, voter files from all three types of vendors were matched to the American Trends Panel, a nationally representative sample of adults who have agreed to take regular surveys. The voter files provide seemingly authoritative data on voter registration and turnout in elections, along with additional demographic, financial, lifestyle and political data. In turn, data from interviews with panelists provides a way to verify the accuracy of some of this ancillary information. Using a common base of matched panel members, voter files from several commercial vendors are compared with respect to a variety of attributes. One analysis of particular interest will focus on whether and how the voter files shed light on the question of why pre-election polls overestimated Hillary Clinton's lead over Donald Trump.

RBS Sampling for Efficient and Accurate Targeting of True Voters

Patrick Ruffini

Election surveys are typically weighted to be demographically representative of the likely electorate, along dimensions like age, gender, race, and educational attainment. Voter file-based sampling allows survey researchers to also weight based on the probability that a respondent will turn out in the election at all, using turnout scores derived from past vote history and informed by basic demographics. This approach addresses the challenges RDD surveys have faced in deriving estimates of likely voter preferences from a sample of registered voters. Public polls in the the 2016 primary and caucus season were vulnerable to criticism for including too many unlikely voters. In Iowa, a record 186,000 Republicans participated in the 2016 caucuses out of a total of 2.1 million registered voters, a 9% incidence rate if using registered voters as a baseline. RDD does not allow pollsters to efficiently target the high-interest universe of likely caucusgoers, but RBS sampling from a voter file or list of past caucus

attendees could. Furthermore, the crucial context provided by a voter's past vote history allows researchers to construct alternative turnout scenarios, by adjusting the weights for the proportion of likely to unlikely voters in the electorate, based on the observed ratio of such voters in previous elections. This approach also has implications for the construction of likely voter models in general elections. Limited data on respondents in the absence of voter files tends to promote all-or-nothing decisions on who is a likely voter, excluding entire classes of respondents who say they are unlikely to vote, when post-election validation finds that many eventually do vote. When used side-by-side with a more traditional approaches, weighting using voter file turnout scores suggested a higher overall rate of undecided voters in the 2016 election and an electorate possibly more up-for-grabs than thought.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

The U.S. Census Bureau's Planning Database as a Free Tool for Survey Practitioners: Applications of the Planning Database in Government, Academia and the Private Sector (Panel)

The U.S. Census Bureau's Planning Database as a Free Tool for Survey Practitioners: Applications of the Planning Database in Government, Academia and the Private Sector

Josh DeLaRosa *Abt Associates*

Matt Jans, Kevin McLaughlin, Joseph Viana, Royce Park, Todd Hughes and Ninez A Ponce

Kathleen Kephart *U.S. Census Bureau*

Matt Jans *University of California - Los Angeles*

Kevin McLaughlin *University of California - Los Angeles*

Tara Becker *University of California - Los Angeles*

Joseph Viana *University of California - Los Angeles*

Royce Park *University of California - Los Angeles*

Todd Hughes *University of California - Los Angeles*

Ninez Ponce *University of California - Los Angeles*

Joe Murphy *RTI International*

Paul Biemer *RTI International*

Darryl Creel *RTI International*

Paul Schroeder *Abt Associates*

Anders Hansen *Abt Associates*

Joseph McMichael *RTI International*

Kevin P Tolliver *U.S. Census Bureau*

Benjamin Reist *U.S. Census Bureau*

To meet the requirement for planning, implementing, and evaluating Census 2000, the U.S. Census Bureau developed the Planning Database (PDB). The PDB contains housing, person and operational data which have informed the U.S. Census Bureau's planning for Census 2000 and the 2010 Census. Beginning in 2012, the U.S. Census Bureau has provided annual updates to the PDB, the latest vintage of which was released in September of 2016. Ever since it was rebooted in 2012, the PDB has contained U.S. housing, demographic, socioeconomic and operational statistics based on select 2010 Decennial Census and select 5-year American Community Survey (ACS) estimates. Data are provided at both the census block group and tract levels of geography. Also since 2012, the PDB has included the Low Response Score (LRS). The LRS is a model-based prediction of mail return rates by block group and tract that may be used separately or in conjunction with other PDB variables. The PDB has been a rich source of open data for researchers within and outside of the U.S. Census Bureau. Potential uses include: • Identifying geographic areas for special outreach and promotional efforts • Examining expected completion rates at low levels of geography • Linking with spatial map data to create thematic maps • Generating reports, cross tabulations, and simple analyses • Planning recruitment activities for interviewer-administered surveys and censuses Data from the PDB has enhanced researchers' ability to develop adaptive data collection algorithms and targeted incentives. With the help of the PDB, researchers have been able to reduce costs across the survey lifecycle while also increasing data quality. This panel will highlight a diversity of PDB use cases. The panel will discuss the utility of the PDB from the perspective of academic, government and private sector uses.

Community v. Household Resistance and Geographic Predictors of Nonresponse Bias: Combining call histories and the Census Planning Database (PDB) in the California Health Interview Survey (CHIS).

Matt Jans, Kevin McLaughlin, Joseph Viana, Royce Park, Todd Hughes and Ninez A Ponce

It is nearly impossible to disentangle the multiply-determined forces that lead to nonresponse and nonresponse bias in real-world surveys. To understand nonresponse bias, ideally, the researcher has data on respondents and nonrespondents. When a sampling frame, as purchased, provides little additional data (e.g., RDD sampling frames) appending auxiliary or other contextual data to the frame can make nonresponse bias analysis possible. This study uses the US Census Bureau's Planning Data Base (PDB) and contact attempt history data from the California Health Interview Survey (CHIS, an RDD survey) to assess the relative roles of community (i.e., geographic) and sample unit causes of nonresponse and nonresponse bias. Using 2011-2014 CHIS data, we measure household-level resistance from call records (e.g., refusals, appointments). Our measure of community-level resistance, the Low Response Score (LRS, part of the PDB), was appended to the RDD sampling frame at the rate center level for all sampled units. Additional auxiliary data were appended from the American Community Survey (ACS) and used as demographic controls. Findings show that the LRS is a useful predictor of screener nonresponse, even when considering other community-level characteristics. However, household-level resistance (measured by refusals and other case dispositions) is much more predictive of final nonresponse and nonresponse bias than community-level resistance. The findings are discussed in the context of surveying hard-to-reach (i.e., hard-to-count/survey) and multilingual and multicultural populations. The approach provides a tool for understanding nonresponse in RDD samples, and other frames with limited data.

Using the Census Planning Database to Tailor a National Mixed-Mode Survey

Joe Murphy, Joe McMichael, Paul Biemer and Darryl Creel

The U.S. Census Bureau's Planning Database is a valuable resource for Census in designing and tailoring studies. The Database also represents a potential value for the external research community. Among other metrics, the Low Response Score provides researchers with the ability to set expectations for mail survey response at areas of geography as small as the block group. This score identifies block groups and tracts whose characteristics predict low census mail return rate and are highly negatively correlated with census and survey participation. We begin by examining the Low Response Score as a predictive measure of mail return and other outcome rates in several national and state surveys that use address-based samples. Next, using the Score, other Database measures, and outcomes from several experimental mode and contact protocols, we simulate the effect of using the Database for tailoring both sampling and outreach strategies in terms of expected response and costs. We focus on outcomes from the 2015 Residential Energy Consumption Survey and Household Pilots, sponsored by the U.S. Energy Information Administration. These surveys contacted households both by mail and in-person. Several experiments were included to inform a design that optimized data quality, cost, and timeliness. Using the Database, we can assess the value that would be added by tailoring the design from the outset to increase sample sizes in areas of low expected response to reduce the effect of unequal weighting that can result when response rates vary across areas. We can also assess the value of tailoring our mode protocols based on correlation between the Low Response Score and survey outcomes. Finally, we can assess whether further outreach strategies may be identified using other demographic measures in the Database. We will provide guidance for other surveys using the Database for planning at the sampling and outreach stages.

Missing in Action: Predicting Item Non-Response for Key Variables

Paul Schroeder, Josh DeLaRosa and Anders Hansen

Missing data or logically inconsistent data can have quality and cost implications on surveys. Without data on key indicators of interest the entire case may be left out of the analysis. The ability to estimate the amount of missing data prior to data collection would allow survey researchers to plan for additional resources needed to mitigate this risk. Building on the methodology to develop the Census low to respond score (LRS), this presentation will explore the feasibility of creating an item-non response score for key indicators such as income, age, and education. Using the overall housing unit characteristic imputation rate as the dependent variable, this research will explore the variable leading to missing data at the tract level.

Constructing an Address-level Low Response Score (ALRS) for Address Based Sampling (ABS) frames.

Joe McMichael and Joe Murphy

The availability and use of auxiliary data continues to grow, particularly for surveys drawn from ABS frames. This trend is allowing survey researchers to develop new tools and methods for improving quality and reducing the costs of their surveys. Inspired by the Low Response Score (LRS) found on the U.S. Census Bureau's Planning Database (PDB), we construct a response propensity score using similar machine learning techniques outlined by Erdman and Bates, 2014. Using the response outcome from a mail survey of 63,000 U.S. residents as the dependent variable and independent variables such as: those commonly found on the USPS Computerized Delivery Sequence (CDS) file; Census Planning Database; Acxiom Infobase marketing data; and other sources, we construct a few variants of an ALRS. We then examine and evaluate how well the new ALRS would have performed on two other mail surveys. This paper discusses how the ALRS models were built, how they performed, and what we could have done differently if this information was available to us during the sample design stage.

Developing Cross-Survey R-indicators Using the U.S. Planning Database

Kevin Tolliver and Benjamin Reist

In this presentation, we discuss how publically available data were used to calculate comparable Representativeness (or R-) indicators (Schouten et al., 2009) for demographic surveys. Declining response rates in demographic surveys (Brick & Williams, 2013) and increasing costs of survey data collections are among the reasons for a general call to move beyond response rates as the sole, near real-time measures of survey data quality. The R-indicator (Schouten et al., 2009; Schouten et al., 2011) is one such measure. It provides information about respondent balance (compared to the originally selected sample) on key auxiliary data. A criticism of the R-indicator is that estimates are only comparable across surveys if a common set of variables is used. The U.S. Planning Database, which includes 5-year American Community Survey block-group level estimates, provides the necessary set of variables to produce comparable R-indicators for surveys that use a geocoded sample.

Discussant

Kathleen Kephart

Kathleen is a member of the Census Planning Database team. Their efforts to maintain and disseminate the Planning Database help 2020 Census stakeholders make data driven decisions. Kathleen will summarize the panel's presentations and provide insights into future PDB enhancements

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

AAPOR's Transparency Initiative (Panel)

AAPOR's Transparency Initiative

Ashley Kirzinger *Kaiser Family Foundation*

Ashley Kirzinger

Timothy Johnson *University of Illinois - Chicago*

Timothy Triplett *The Urban Institute*

Mandy Sha *RTI International*

Mikelyn Meyers *US Census Bureau*

Natalie Jackson *Huffington Post / POLLSTER.COM*

Robin Kaplan *Bureau of Labor Statistics*

Jennifer Edgar *Bureau of Labor Statistics*

With more than 80 organizations, the growing responsibilities of the Transparency Initiative includes taking a step back to assess whether transparency, and a consistent disclosure of methods, leads to a better polling environment. In an effort to answer this very broad and complex question, members of the Transparency Initiative Coordinating Committee have organized a panel examining various issues around transparency in polling, including an overview of the development of AAPOR's Transparency Initiative, the role of transparency in the 2016 presidential election horserace polling, as well as looking at ways of going forward in expanding transparency to qualitative polling methods, and what's next for the TI.

Looking Back: How AAPOR Advocated for and Implemented Disclosure Requirements

Ashley Kirzinger

Since 1947, pioneers of public opinion research, including George Gallup and Stuart Dodd among others, have been advocating for higher levels of disclosure among survey research organizations. In 1992, Frank Newport argued that polling results need to replicate the scientific and scholarly model such that "the manner in which most polling results are in fact reported to the American public doesn't conform to these requirements or objectives." In 2010, AAPOR launched the Transparency Initiative in an effort to encourage broader and more effective disclosure of research methods by all organizations. The goal of the TI is to promote methodological disclosure through a proactive, educational approach that assists survey organizations in developing simple and efficient means for routinely disclosing the research methods associated with their publicly-released studies. Before we go forward and examine the current state of the TI and what's next, we will take a look back at the development of AAPOR's Transparency Initiative and why the organization advocated for disclosure standards in public opinion research.

Transparency and the 2016 Polling

Natalie Jackson

In 2012, just 9 percent of election pollsters reported all of AAPOR's survey disclosure items in their results. This study looks at the 2016 election polling and examines whether organizations disclosed more information in their polling results and whether disclosure practices vary among organizations by affiliation (AAPOR membership, academic, partisan) and methods (probability and non-probability).

Transparency and Quality in Multilingual Cognitive Testing

AAPOR Transparency Initiative requires 12 items to be disclosed in the reporting of qualitative research. In theory, proper disclosure of methodological information that is consistent with these standards should enable any researcher to evaluate for themselves whether a particular study met standards of quality. In the case of multilingual cognitive testing studies, we are not aware that this approach has been put into practice. Two challenges are that the researcher may not speak multiple languages and it may be hard to systematically examine the disclosed information when “tailoring” is often necessary to accommodate cultural and linguistic variation. Are these transparency standards relevant to multilingual cognitive testing and its quality? To answer this research question, we used AAPOR disclosure items to reformat a report of Chinese, Korean, Vietnamese, Arabic, and Russian cognitive testing. We then compared the reformatted report’s completeness with the reporting that applies the two quality frameworks that could be utilized in multilingual cognitive testing: (a) Boeije and Willis (2013) developed the Cognitive Interviewing Reporting Framework (CIRF), which uses a checklist approach to allow systematic reporting of the methods and includes 10 criteria; and (b) The Roller and Lavrakas (2015) Total Quality Framework (TQF) applies four criteria, including Transparency, to maximize quality outcome. We found that the CIRF checklist format was effective in reminding us of the information to include to enable systematic evaluation, while the TQF could adapt to the unique considerations of multilingual cognitive testing and connects with other qualitative methods. Results from both frameworks support all 12 AAPOR disclosure items. Our findings suggest that transparency is directly relevant to evaluating the quality of multilingual cognitive testing, especially when the researcher does not speak multiple languages. AAPOR guidelines provide the standards for transparency, and researchers should adopt the most suited framework for the specific studies.

The Future of Transparency in an Uncertain Age for Polling

Tim Triplett

If the 2016 election taught us anything, it was there is strong demand for transparency in the polls that are being conducted and more broadly for surveys as a whole. The chatter before and particularly after the election often talked about the need for more organizations to become transparent and for more rigorous transparency requirements. There was interesting dialogue that included having transparency ratings and requirements that would make it mandatory for organizations to share all data collected among other suggestions. This presentation will focus on having open discussion about the feasibility of many of the transparency issues that were raised in light of the 2016 election polling and how they could help shape the future of the Transparency Initiative. Keeping in mind, the AAPOR transparency Initiative started accepting member in October of 2014 and many of members are now receiving their biennial review. So, the discussion will also need to consider how any changes to the Transparency Initiative could impact current members.

Confidentiality Concerns, Do They Matter More than Confidentiality Pledges?

Robin Kaplan

Protecting respondent confidentiality is a top concern for survey organizations, and determining how that protection is communicated to respondents can be a challenge. Ideally, confidentiality pledges are easy to understand and reassure respondents that their data are kept confidential, while not raising additional concerns, which may negatively impact response rates (Couper et al., 2008). Many government surveys have used the same language for many years, however, the passage of the

Cybersecurity Enhancement Act of 2015 resulted in a change to the way data are protected, requiring revisions to the confidentiality pledges. To better understand the impact of the revised confidentiality pledge, the Bureau of Labor Statistics (BLS) conducted research on people's comprehension of, and reaction to, both the current and revised pledge. A total of 864 participants were recruited from Amazon's Mechanical Turk to complete an online survey. They were randomly assigned to read the current or revised confidentiality pledge. The survey content was designed to gauge initial attitudes and perceptions about data confidentiality, as well as comprehension of and reactions to the confidentiality pledge. Despite the reported concern for data confidentiality, we found evidence that participants did not seem to focus on either version of the pledge, with very few differences in comprehension or reactions by pledge version. Additionally, item non-response and attrition rates did not differ by pledge version. However, the level of concern about confidentiality participants reported led to group differences in several data quality metrics (attrition, item non-response, time spent reading the pledges, straightlining), as well as in comprehension of the pledge. This suggests that perceptions and attitudes toward data confidentiality may affect survey responses and data quality more than the actual language used to communicate it. This presentation will summarize the results of this study, discuss potential implications, and follow-up research.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Methodological Brief: Using Incentives to Increase Survey Participation and Reduce Costs

Taking Chances: Are Sweepstakes an Effective Incentive for Compliance?

Arianne Buckley *Nielsen*

Erin Wittkowski *Nielsen Company*

Incentives are commonly used to improve cooperation. When deciding upon incentives, researchers must consider budget, cooperation rate goals, and level of respondent burden. Is the compliance gained worth the cost of the incentive? Also, is the incentive being offered enough to motivate respondents to cooperate? Nielsen participants receive weekly incentives in return for their cooperation. At times of the year when sample performance decreases, panelists are offered special one-time incentives for meeting specific compliance thresholds. These special incentives take the form of either a specified monetary payout or entries into a large monetary sweepstakes. The majority are a guaranteed monetary amount, but while they're successful at improving compliance, can become costly. The sweepstakes incentive, if won, is a much larger payout for panelists but less expensive overall. Given this, the sweepstakes incentive is an attractive option for incentivizing desired behavior if it is as effective as other incentives. In December 2016, we will conduct a test to re-examine the effectiveness of the sweepstakes as a cost-effective method for incentivizing panelists. The test is designed to answer two research questions. First, are sweepstakes with large monetary rewards effective at improving performance? Second, are sweepstakes as impactful as offering a much smaller but guaranteed amount of money for meeting a specific compliance goal? For this test, panelists will be divided into two groups: (1) a test group where each panelist receives entries into a sweepstakes each day they meet their compliance goals and (2) a control group where panelists are not offered the sweepstakes. The analysis will compare panelist compliance rates between the test and control groups. The compliance rates from this test will also be compared to compliance rates from previous bonus incentives. The results will help determine if incentives of chance are an impactful and cost-effective option for improving compliance.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Methodological Brief: Using Incentives to Increase Survey Participation and Reduce Costs

Incentive Visibility in a Mail Survey of Physicians

Marshica S. Kurtz *RTI International*

Emily M. Geisen *RTI International*

Rebecca J. Powell *RTI International*

Joe J. Murphy *RTI International*

Murrey G. Olmsted *RTI International*

To maintain high response rates, researchers often use incentives to motivate response. For mail surveys, Dillman (2007) recommends ensuring that incentives can be seen immediately upon opening a letter. For example, cash incentives should be placed on top of the questionnaire and folded inside of the cover letter so that all survey materials are removed from the envelope at the same time. Yet, preparing incentives in this way precludes the use of automated folding and inserting, increasing the costs of the mailings. Aside from Dillman's suggestions, we could find no recent empirical research that investigates the purposeful positioning of incentives for mailed surveys. In this research we investigate the effect of increasing incentive visibility on mail survey response rates with physicians. Our study tests the recommendations by Dillman (2007) on a representative sample of 4,700 physicians from the 2017 Best Hospitals Physician Survey. As an incentive, we used a \$2 bill and a pen, because previous research found higher response rates for physicians who received both compared to those who received only a pen or only a \$2 bill (Kurtz et al., 2016). We randomly assigned the physicians to one of two conditions: (1) the pen is used to clip the \$2 bill to the top of the letter and (2) the pen and \$2 bill placed in the envelope with no secure positioning. For both conditions, bubble envelopes were used to ensure that the pens did not tear the envelopes. We compare response rates, timing of responses, differences in survey estimates, and the cost of the two groups at the end of data collection. The implications of this work may provide researchers with additional techniques to boost the influence of incentives.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Methodological Brief: Using Incentives to Increase Survey Participation and Reduce Costs

The effectiveness of incentives on recruitment and retention rates: an experiment in a web panel

Salima Douhou *City, University of London*

Annette Scherpenzeel *Munich Center for the Economics of Aging*

Joris Mulder *CentERdata, Tilburg University*

An important instrument to secure respondent cooperation and the quality of responses is the use of (monetary) incentives. The purpose of this paper will be to examine which incentive level is related to long term participation of respondents. An experiment was carried out during the recruitment of a refreshment sample of the LISS panel in 2011. The LISS panel is a probability-based online panel and panel members are invited every month for completing questionnaires for which they are paid an incentive of 15 euros per hour. A random half (~1000 households) of the refreshment sample was, in the recruitment stage, promised a higher monetary payment for completing questionnaires (25 euros per hour). We are interested in the relation between the incentive levels and signing up for participation and the relation between the hourly incentive level (15 and 25 euros) and long term participation in the LISS panel. To study this we will model the time-to-event of attrition using survival analysis, which has the advantage that it incorporates the timing of the event and it allows for censoring. We find that the higher incentives can be used to stimulate longer participation of younger participants and that respondents in the high incentive group perform better in terms of retention.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Methodological Brief: Using Incentives to Increase Survey Participation and Reduce Costs

Who accepts payment when remuneration is offered?

Alisha Baines Simon *Minnesota Department of Health - Health Economics Program*

Sarah LHagge *Minnesota Department of Health - Health Economics Program*

Kathleen Thiede *Univ. of Minnesota, SHADAC*

Kendal Orgera *Minnesota Department of Health - Health Economics Program*

Giovann Alarcon *University of Minnesota, SHADAC*

Karen A Turner *University of Minnesota, SHADAC*

Extensive research has explored the impact of survey incentives on response rates and sample bias. However, there is little research on who accepts remuneration for cell phone minutes. The majority of cell phone plans are unlimited, where the customer pays a monthly fee instead of a minute-by-minute rate. However, there is still a percentage of people who have prepaid cell phones. Without remuneration for those minutes, respondents incur cost for survey participation. Every two years, the state of Minnesota conducts the Minnesota Health Access Survey (MNHA), a population-based dual frame telephone survey on health insurance coverage and access. In 2015, the MNHA sampled prepaid cell phones at a 2 to 1 ratio to non-prepaid cell phones in order to increase the likelihood of reaching certain populations. In order to off-set the cost of cell phone minutes, a remuneration of \$5 was offered to all cell phone respondents. This study explores characteristics of those who accepted the remuneration. First, we explore how those who accepted remuneration differed from those who did not in terms of demographics and health insurance status. Though those with prepaid cell phone plans were more likely to accept, cell phone respondents accepted remuneration even when they were not on prepaid plans. About 62% of the prepaid cell phone sample accepted the remuneration: a little more than half of the non-prepaid sample accepted. Remuneration acceptance rates also differed by other demographic characteristics. For example, take-up rates were about ten percentage points higher for those with incomes under 100% federal poverty guidelines (FPG) as compared to those with incomes over 400% FPG. Second, given higher uptake rates among prepaid cell phone respondents and other subgroups, we simulate how the demographics of our sample may have changed if the remuneration had not been offered.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Methodological Brief: Using Incentives to Increase Survey Participation and Reduce Costs

Experimenting with Incentive Amounts and Structures to Maximize Effectiveness and Efficiency

Kristen Faucetta *MDRC*

Charles Michalopoulos *MDRC*

Lee Robeson *Survey Management, Inc.*

Eileen Bandel *MPR*

Martha Kovac *MPR*

Erin Panzarella *MPR*

David DesRoches *MPR*

While incentives are a tool often used to encourage participation in surveys and other data collection activities, it is unclear what incentive level and structure will maximize response rates efficiently. This presentation will discuss results of two experiments conducted within MIHOPE (Mother and Infant Home Visiting Program Evaluation) of different incentive schemes and amounts. The goal of the experiments is to learn which incentives are most effective and most efficient for the MIHOPE sample (young, mobile, low-income, racially and ethnically diverse mothers of young children). Using a factorial design, the first experiment randomized 1,220 study participants into six experimental conditions to test varying incentive levels for two components of data collection (a phone survey and a set of in-home assessments) and offering an early bird incentive (a higher incentive for response within the first few weeks of data collection). The second experiment also used a factorial design to randomize 1,705 study participants into four experimental conditions to test two different incentive structures for a short web/phone survey: an early bird incentive and a partially prepaid incentive. Factorial designs allow the main effects of the incentives to be estimated using the full sample and to examine interactions between the different incentives. Analyses will look at both efficacy (examining whether there are statistically significant differences in response rates across experimental groups at the conclusion of the data collection period and at different numbers of weeks after data collection launch) and efficiency (comparing the level of effort needed to achieve response rates across experimental groups). These analyses will be an important contribution to the limited experimental evidence (particularly for low-income samples) around incentive structures and amounts.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Methodological Brief: Using Incentives to Increase Survey Participation and Reduce Costs

Effects of Sequential Prepaid Incentives to Increase Participation and Data Quality in a Mail Survey of Pediatricians

John R Stevenson *Univ. of Wisconsin, Survey Ctr.*

Jennifer Dykema *Univ of Wisconsin Survey Ctr*

Chad Kniss *Univ of Wisconsin Survey Ctr*

Nadia Assad *Univ of Wisconsin Survey Ctr*

Cathy Taylor *Tulane University*

While collecting data that are of high quality in surveys of physicians is an important goal, response rates in physician surveys are often low. Several studies find that the inclusion of a single, small, prepaid monetary incentive is effective in increasing response rates among physicians. Little research, however, examines what effect variation in the distribution and administration of these incentives has on response rates, and recent recommendations regarding incentives advise researchers to include a second cash incentive in a follow-up contact in order to increase the likelihood that “later communications will be read, and hopefully acted upon (Dillman et al. 2014, 424).” We embedded an experiment in a mail survey to examine the effects of different incentive combinations on participation. The sample for this study consisted of 1,500 randomly selected pediatricians, purchased from a private vendor. In early 2016, sample members were randomly assigned to the following incentive groups: \$10 delivered in the first mailing only; \$5 in the first mailing/\$5 in the second mailing; \$5 in the first mailing/\$10 in the second mailing; \$10 in the first mailing/\$5 in the second mailing; and \$10 in the first mailing/\$10 in the second mailing. We examine which of the following strategies was most effective: (1) providing the full amount in the initial mailing and not offering a second incentive versus (2) splitting the full amount equally between the first and second mailings versus (3) splitting the full amount and providing more in the first mailing and less in the second mailing versus (4) splitting the full amount and provide less in the first mailing and more in the second mailing. In addition to examining differences in response rates among the groups, we also conduct an analysis of how the groups varied in terms of costs, demographic representativeness, and item-nonresponse.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Methodological Brief: Using Incentives to Increase Survey Participation and Reduce Costs

Improving General Population Survey Response Rates with Visible Money

Matthew DeBell *Stanford Univ.*

Natalya Maisel *Stanford University*

Brad Edwards *Westat*

Michelle Amsbary *Westat*

Vanessa Meldener *Westat*

In mail surveys, and in advance letters for surveys in other modes, it is common to include a prepaid incentive of a small amount such as \$1, \$2, or \$5. However, when letters are addressed to "Resident" or "The family living at" the sampled address, advance letters may be perceived as junk mail and thrown away without being opened, so the enclosed cash is wasted. In the current study we describe the results of experiments testing new ways of letting respondents know that their letter contains cash. The study uses address-based sampling to randomly select households in the US. We test an advance postcard announcing the impending arrival of a letter with enclosed cash. We also test the novel use of a window envelope in which the enclosed cash is clearly visible from outside the sealed envelope, and we test the US Postal Service for evidence of rates of theft of such letters. The study has the potential to demonstrate low- or nearly zero-cost methods to improve response rates in mail surveys and in surveys in other modes that use advance letters containing prepaid incentives.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Interviewer Behaviors, Performance and Effects

Estimating Interviewer Effects in the Absence of Interpenetration

Michael Elliott *University of Michigan*

Brady West *Univ of Michigan, ISR*

Interviewer effects result from responses collected by the same interviewer being correlated, reducing the efficiency of survey estimates and decreasing effective sample sizes. While having received attention in early survey methods research, this well-known problem has been much less studied recently. This may be due to the fact that standard approaches for assessing interviewer effects requiring random assignment of cases to interviewers (interpenetrated sampling), which is often infeasible in many sample designs. In the absence of interpenetration, standard clustering methods may overestimate interviewer effects by confounding non-random assignments of respondents with measurement error introduced by interviewers. Here we propose two methods to assess interviewer effects in the absence of interpenetration: “anchoring,” which uses variables assumed to have little or no interviewer measurement error to remove intra-interviewer variability due to assignment via joint distribution modeling assumptions; and “propensity adjustment,” which uses weights developed from interviewer assignment propensity models to provide pseudo-interpenetrated-design estimates.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Interviewer Behaviors, Performance and Effects

The Impact of Interviewer Effects on Regression Coefficients

Brady T West *Univ of Michigan, ISR*

Micha Fischer *University of Michigan*

Michael Elliott *University of Michigan*

Frauke Kreuter *Univ. of Maryland, JPSM*

This presentation will examine the influence of interviewers on the estimation of regression coefficients from survey data. First, we will present theoretical considerations with a focus on measurement errors and nonresponse errors due to interviewers. Next, we will present the results of a simulation study identifying which of several nonresponse and measurement error scenarios has the biggest impact on the estimate of a slope parameter from a simple linear regression model. We find that when response propensity depends on the dependent variable in a linear regression model, bias in the estimated slope parameter is introduced, but we also find no evidence that interviewer effects on the response propensity have a large impact on the estimated regression parameters independent of the missing data mechanism. The simulation study also suggests that standard measurement error adjustments using the reliability ratio (the ratio of the measurement-error-free variance to the observed variance with measurement error) can correct most of the bias introduced by interviewer effects in a variety of complex settings, indicating that more routine adjustment for such effects should be considered in regression analysis using survey data. Finally, we examine the primary sources of interviewer effects on regression coefficients estimated using real survey data collected in Germany, in a study where the nonresponse errors and measurement errors introduced by interviewers for individual variables could be computed using administrative data. We find that our proposed adjustment works well to correct most of the bias introduced in this setting.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Enhancing the Quality of Internet Data Collection

Moving Establishment Surveys from Mail to Web: Performance and Data Quality

Georg-Christoph Haas *Institute for Employment Research*

Stephanie Eckman *RTI International*

Ruben Bach *Institute for Employment Research (IAB)*

Frauke Kreuter *Univ. of Maryland, JPSM*

Moving existing establishment surveys from face to face, mail and telephone to the web can make data collection cheaper and more efficient and allow the use of techniques like dependent interviewing, plausibility checks and complex filters. However, there is little research into the effect of switching to web data collection on data quality and performance rates. We conducted a survey of establishments in Germany which experimentally varied five mode conditions: web only, mail only, web/mail choice condition, mail with web follow up and web with mail follow up.. By comparing responses to administrative records, we can report the measurement error in each condition. We also compare the unit and item nonresponse rates across the conditions. We also explore whether the best mode varies by establishment size and/or industry. Our results are relevant for establishment surveys that wish to move to the web, but are worried about measurement error and nonresponse effects on the data.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

90210 Isn't Hollywood! Or Is It? Issues Related to Geographic Specificity of Cell Phone Samples

Using De-Identified Administrative Records as a Sampling Frame for a Child Health Telephone Survey: Results from a Pilot Study in New York City

Michael Sanderson *NYC Dept of Health and Mental Hygiene*
Stephen Immerwahr *NYC Dept of Health and Mental Hygiene*
Dina Shapiro-Luft *NYC Dept of Health and Mental Hygiene*
Katharine McVeigh *NYC Dept of Health and Mental Hygiene*
Rachel Martonik *Abt Associates*
Nicole Lee *Abt Associates*
Margaret Tyson *Abt Associates*
Andrew Burkey *Abt Associates*
Michael Battaglia *Battaglia Consulting Group, LLC*

The New York City (NYC) Department of Health and Mental Hygiene (DOHMH) plans to conduct a telephone survey of households with children age 0-13, oversampling those with younger children (age 0-5) in specific neighborhoods. Only 23% of NYC households have children in this age group and neighborhood oversampling of households of younger children will be as great as a factor of 3. The cost of screening using a random digit dialing (RDD) telephone sample would be prohibitive. We tested three non-RDD samples designed to efficiently reach households with children and to do so with geographic specificity. These were 1) birth certificates from mothers giving birth in NYC in the past 6 years, 2) contact information for NYC public school children ages 5-13, and 3) a proprietary sample of cell telephone numbers of households believed to have at least one child under age 16. The de-identified administrative records (telephone numbers and ZIP codes) were used only to contact households with children and then randomly sample a focus child for the survey, not to sample a specific child. We evaluated efficiency (working number rates, screening rates, and eligibility rates by child age groups) and geographic accuracy, and compared child and household demographics from completed surveys to Census and DOHMH survey data to assess potential bias. The two NYC administrative list samples were more efficient than the targeted cell sample in reaching households with children age 0-13. Geographic specificity was high (85% or higher) for interviews from all three samples. Child and household demographics of interviews from the birth certificate sample were most similar to Census and DOHMH survey data, while interviews from the targeted cell sample were the least similar. The results confirmed the utility of de-identified administrative records as a sample frame for a health survey focused on NYC children.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Strategies for Improving Instruments and Increasing Data Quality

Enhancing Data Quality Using (Expenditure) Records

Safia Abdirizak *Bureau of Labor Statistics*

Brett McBride *BLS*

Yezzi Angi Lee *Bureau of Labor Statistics*

The Bureau of Labor Statistics fielded the Proof of Concept (POC) test in 2015 to assess methodological, operational, and experiential issues related to the Consumer Expenditure Survey's (CE) redesign plan. The test included one wave of the redesign plan – a recall interview and records-based interview (both having three month reference periods) and one week of diary-keeping for eligible household members. The focus of this presentation is on the records-based interview, which asked questions about expenditures for which respondents were more likely to have records or that were harder to recall for a 3 month reference-period (e.g. cellphone bills, co-pays for medical expenses, and pay stubs). The POC test encouraged the use of records by providing an additional \$20 incentive and instructing respondents to collect records during the recall-based interview the previous week. Record use is an integral part of the redesign plan because past research has shown an association of record use with increased expenditure reporting (Edgar, 2010; Dillman and House, 2012; Davis et al., 2013). We will present our research objectives, which sought to determine whether and how record use influenced the quality of the data that were collected in the POC. Specifically, with the use of records, was there less evidence of rounding to multiples of \$25? Was there less item non-response, whether in the form of don't knows or refusals? Did record use cue additional expenditure reports? And finally, did respondents report increased expenditure totals in certain categories when referencing records? This presentation will be of interest to survey methodologists with an interest in using records to improve response accuracy.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Data Collection in an International Context: Lessons Learned from the Field
Comparison Between the Google Survey and Landline RDD in Two Japanese Regional Elections

Midoriko Nagasaki *The Asahi Shimbun*

Yasuyuki Saito *The Asahi Shimbun*

Nicolaos E. Synodinos *Univ. of Hawaii - Manoa*

The Japanese telephone numbering system makes regional cell phone RDD surveys impractical. Landline RDD is used for such surveys with the inevitable coverage problems that have been increasing. A new method of collecting survey data is the Google Survey (GS). The GS presents pop-up questions to randomly selected individuals that use Google's search engine and are induced to participate by access to contents from various publishers. In the 2012 U.S. presidential election, results from pre-election surveys using the GS were relatively close to the election results and their accuracy was comparable to traditional RDD surveys. The GS was launched in Japan in 2015 and can be used for both nationwide and regional surveys. Asahi Shimbun investigated the accuracy of the GS vis-à-vis landline RDD surveys in two prefectural elections: The 2016 gubernatorial elections in Tokyo and Niigata. For these elections, Asahi conducted surveys using the GS at approximately the same time as its landline RDD surveys. The design and administration of the GS for these two surveys will be presented and differences in their implementation in the two prefectures will be discussed. The results from the GS and the landline RDD surveys were compared with the election results. For Tokyo, the percentage of the vote predicted for the winner with the landline RDD survey coincided exactly with the election results, whereas the results of the GS differed by 2 percentage points. Results of the landline RDD were within 1 to 2 points for the other two major candidates and those of the GS within 3 to 4 points. For Niigata, the winner's percentage differed from the surveys by 12 points (for the landline RDD) and 5 points (for the GS), with both surveys predicting the wrong winner. Additional studies are needed to investigate further the predictive accuracy of the GS.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Enhancing the Quality of Internet Data Collection

Interviewer Gender and Survey Responses: The Effects of Humanizing Cues Variations

Wojciech Jablonski *Utrecht University*

Katarzyna Grzeszkiewicz-Radulska *University of Lodz*

Aneta Krzewinska *University of Lodz*

According to social interface theory (Nass et al. 1996; Nass et al. 1997; Fogg & Nass 1996), humanizing information sent by the computer causes reactions typical of human-to-human interactions. Based on these findings, several methodological research studies have been conducted in order to investigate whether the regularities described by the psychologists can be observed in internet surveys in the form, among other things, of interviewer effect (e.g., Tourangeau et al. 2003; Couper et al. 2003; Fuchs 2009). These research results, however, seem to be inconsistent. In this presentation, we report selected results from an experiment conducted between December 2016 and March 2017 among university students (N=900) as part of the research project funded by the Polish National Science Center. This project aims to estimate the influence of humanizing cues on the quality of the data obtained in internet surveys. The presentation shows the findings concerning the impact of the interviewer's gender on the data. The experiment was based on the multifactorial plan in equal, completely randomized groups. Form of imitation/presence of the interviewer was our major independent variable (factor A) in four scenarios: (1) CAWI/text (with all stimuli presented in the form of text); (2) CAWI/photo (with stimuli presented in the form of text and an interviewer photo); (3) CAWI/movie (with all stimuli presented in the form of video of real interviewers and, additionally, the answers presented in the form of text); and (4) CAPI. The other independent variables were: interviewer gender (factor B, nested within factor A), with two values (male/female) and an extra version with no information for interviewer gender ("we"); interviewer (factor C, random factor nested within factors A & B), where five male and five female interviewers were engaged; and interviewer gender (factor D, constant factor) with two values (male/female).

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

90210 Isn't Hollywood! Or Is It? Issues Related to Geographic Specificity of Cell Phone Samples

Sampling Out-of-Area Numbers for Local Area Cell Phone Samples

Randal ZuWallack *ICF International*

Kristie Healey *ICF International*

Brian Orleans *ICF International*

Melissa Cidade *ICF International*

Scott Heemann *ICF International*

Advances in targeting cell phone samples have increased the efficiency of local area cell phone surveys. Rate centers and billing zip code matching have enabled samples to narrow in on the target geographic area. A typical sampling strategy is to select a random digit dialing (RDD) sample from cell phone banks originally assigned to rate centers in and around the target geography. The sample of cell phone numbers is then matched to identify the zip code of the billing address for the cell phone number. Numbers linked to a zip codes inside the geography can be oversampled, while numbers linked to a zip code outside the area can be excluded or undersampled. One limitation that has hampered this two-phase sampling design is the inability to reach residents of the target area with a cell phone number associated with a different area. Until recently, there was no feasible way to reach out-of-area numbers, thus local area samples were at risk for coverage bias. Survey Sampling International's (SSI) Smart CellTM now offers the capability to sample from a segment of out-of-area numbers no matter where they originated. Our first experience with out-of-area sample was for Santa Barbara County, CA, where we selected a sample of cell phone numbers with a billing location in Santa Barbara County, but did not originate from the Santa Barbara rate centers. Forty-one percent of the out-of-area respondents reported that they lived in Santa Barbara County, lower geographic eligibility than the numbers selected from the Santa Barbara rate centers (78%). We are now conducting out-of-area samples in Vermont, Houston, San Francisco, and San Jose. Our presentation reports the accuracy results for these areas; compares the demographics of the out-of-area respondents with those of the in-area respondents; and discusses sample stratification and allocation considerations.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Strategies for Improving Instruments and Increasing Data Quality

The Impact of Respondents Changing Previously-Collected Data in Subsequent Rounds of Longitudinal Surveys: an Application to the MCBS

Megan H Stead *NORC at the University of Chicago*

Lauren McNamara *NORC at the University of Chicago*

Jennifer Vanicek *NORC at the University of Chicago*

Nicholas Schluterman *Centers for Medicare & Medicaid Services*

Joseph Regan *Centers for Medicare & Medicaid Services*

The Medicare Current Beneficiary Survey (MCBS) is a continuous, multipurpose survey of a nationally representative sample of the Medicare population, conducted by the Centers for Medicare & Medicaid Services (CMS) through a contract with NORC at the University of Chicago. The MCBS is a longitudinal panel survey that interviews participants three times per year for four years, with each interview termed a “round”. A unique feature of the MCBS interview is “Summary Sections” which allow the interviewer to review responses with the respondent to confirm or change responses from the prior round interview for health insurance plans and prescribed medicines. Changing data in a Summary Section rests on the assumption that a change is needed and that recall during the current interview is more accurate than recall during the prior interview. This paper will focus on the Health Insurance Summary section of the MCBS, reviewing multiple rounds of data to determine the extent to which respondents change prior-round responses. First, this paper will review the data collection intentions and survey process. Second, data analysis will review what types of information is changed, look for patterns in the changed information, and measure how much the changed information differs from the previously-collected data. Additional analysis will segment by health insurance plan type looking for further insight. By studying this unique feature of the MCBS, this research aims to better understand whether Summary sections are an insightful tool for improving data quality.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Strategies for Improving Instruments and Increasing Data Quality

Integrating Feedback from Post-Collection Data Management into Questionnaire Design

Reina Sprankle *Westat - Rockville, MD*

Beth Hunter *Westat*

Brandi McMillan *Westat*

Jocelyn Newsome *Westat*

Jennifer Anderson McNulty *Westat*

Kerry Levin *Westat*

Brenda Schafer *Internal Revenue Service*

Patrick Langetieg *Internal Revenue Service*

Saurabh Datta *Internal Revenue Service*

Research has shown the importance of visual design in ensuring respondents correctly complete self-administered questionnaires (Dillman et al, 2014; Millar and Dillman, 2011). Expert reviews, cognitive testing, and pilot studies can identify some design limitations prior to data collection. For ongoing studies, an additional source of invaluable feedback can come from data management staff. As the front line for viewing and processing data, staff can provide valuable information about overlooked instructions or misleading visual cues. These design issues can result in respondents inaccurately filling out questionnaires, impacting data quality. We will examine the role of data management feedback in enhancing the design, quality and efficiency of survey instruments for a large annual household survey. The IRS Individual Taxpayer Burden (ITB) Survey is an annual multi-mode survey sent to 20,000 individuals in the United States. It measures the time and money taxpayers spend complying with tax law. The IRS ITB Survey is currently being fielded for the sixth consecutive year. With each fielding, data managers examined the post-collection data processing operations for areas where respondent mistakes prevented automatic data capture and required manual intervention. Data managers identified issues related to the design and setup of the surveys by reviewing marginal comments, missing data, misplaced data, and other anomalies. Some examples include respondents confusing a comma with a decimal when entering dollars and cents; writing in their own response options for an education demographic item; and overlooking a “not applicable” response option due to its placement beneath a large “other specify” box. Based on these findings, researchers revised items for subsequent fieldings of the ITB Survey. We will discuss how data processing feedback was successfully integrated into the questionnaire design process. We will also examine the impact that these modifications had on data capture quality and efficiency.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

90210 Isn't Hollywood! Or Is It? Issues Related to Geographic Specificity of Cell Phone Samples

Hitting Them Where They Live: Comparing List-Assisted RDD Cell Phone Sample to Rate Center Based RDD Cell Phone Sample

Missy Mosher *SSI*

Linda Piekarski *SSI*

Randal ZuWallack *ICF*

From the time that researchers began generating RDD cell phone sample, drawing the sample from rate centers located in and around their target geography was the best option available. This geographic method has frequently proven itself to be imprecise. There are issues of both over-coverage (calling numbers within a rate center that connect with people who live outside the target geography) and under-coverage (missing residents with numbers belonging to a rate center outside the sampled geography). Using the billing Zip Codes from over 220 million phone numbers in the US, a 'list-assisted' RDD cell phone frame can be built. As with traditional RDD list-assisted landline frames, the sample would be a proportionate blend of 'listed' phone numbers, where the billing address is known to match the target geography (FIPS Code and/or ZIP Code), and unknown or 'unlisted' numbers. Business suppression may also be utilized in this approach. Results from dialing a 'list-assisted' RDD cell sample side-by-side with a traditional rate center based RDD cell sample will attempt to measure any reduction in over-coverage and/or under-coverage, any increases in efficiency and any improvement in demographic distribution as compared to Census values for the area when using this frame. If successful, the creation of a list-assisted RDD cell phone frame could give researchers a more representative and efficient alternative to traditional rate-center based RDD frames.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

90210 Isn't Hollywood! Or Is It? Issues Related to Geographic Specificity of Cell Phone Samples

Evaluating coverage bias and efficiency in RDD telephone surveys under different frame constructions

Burton Levine *RTI*

Joel Hampton *RTI*

Dual frame surveys, utilizing a frame of landline phone numbers and a frame of cell phone numbers, have become the norm for telephone surveys. Some survey researchers have suggested sampling exclusively from a cell phone frame to reduce cost with a limited increase in bias. Another alternative is sampling from the cell phone frame and the listed landline frame, but excluding unlisted landline frame members. Compared to excluding the entire landline frame, excluding only the unlisted frame will capture all the cost savings and result in potentially less coverage error. The increase in coverage bias is investigated in the two exclusion scenarios for various health outcomes using the 2015 Behavioral Risk Factors Surveillance System public use data. For each state and the District of Columbia, separately, two alternative respondent populations are formed: (1) cellphone respondents, and (2) cellphone plus listed landline respondents. For each alternative respondent population, the design weights are recalibrated in order to create new analysis weights that are representative of the target population (i.e., each state). Various estimates of health outcomes are calculated using all the full respondent population and the two alternative respondent populations. Bias, using the full respondent population estimate as the true measure, and the loss of precision due to the unequal weighting effect (UWEs) are calculated for each of the alternative respondent populations. Finally, data collection costs, for each respondent population are evaluated to determine if one of the alternative respondent populations provide significant cost savings while only minimally increasing coverage bias.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Data Collection in an International Context: Lessons Learned from the Field
Problems of surveying public opinions in the Arab world as seen by the academic elite: a field study
Reda Abdelwaged Yousef *Ahlia University*
Hemat Alsaka *Ahlia University*

Abstract Surveying public opinions faces many problems, especially in the Arab World, as people seem to devalue the importance of honest responses to such surveys. Additionally, there are very few reliable survey and measurement centers in the Arab region despite the scientific value and contribution that surveys can have for understanding the different communicative, social, cultural and political phenomena. There have been limited surveys of public opinions in the Arab world which could be caused by a number of problems and difficulties. This study tries to identify these problems and difficulties through a field study. The main question of the current study is: What are the main problems facing the measurement of public opinions in the Arab world from the perspective of the academic elite? The study also tries to identify the difficulties faced by researchers and research centers during their surveys of the public opinions on different issues and phenomena. The study is descriptive in nature and uses a survey as a methodology. It attempts to provide a scientific description of why there are scarce and reliable surveys measuring public opinions in the Arab world and tries to find out the reasons that led to this phenomenon. The study uses a questionnaire as tool to collect data from the participants. To make sure that it is suitable for achieving the study objectives, the questionnaire has been piloted on a random sample of academics in the social sciences representing the east region of the Arab world such as the Gulf states and the east region such as Algeria, Morocco, Tunisia, as well as the countries of the center of the Arab World, such as Egypt and Sudan. Key words: public opinions, measurement of public opinions, the Arab World, the elite

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Issues in Opinion Formation and the Media

Just-in-time Information Acquisition: Empirical models from health and science

Jon Miller *University of Michigan*

Human information acquisition is undergoing a fundamental change and it is important that social scientists engaged in survey measurement to understand the nature of these changes and the implications for learning, communication, and education in modern societies. For most of human history, human information acquisition (an umbrella term encompassing education, communication, and inter-personal conversation and discourse) has been based on a warehouse model that involves the provision of information from a more learned person, a political or religious leader, or media sources and the individual was expected to store the new information in his or her mental warehouse for future use. Like all warehouse systems, this mode is in the initial stages of failure and replacement. In the electronic era, individuals seek information on topics that they want or need to know about and retain the information that they think useful. The rapid expansion of the internet and wireless communication has facilitated this process and new and emerging technologies will expand it. The new information market place is more like a bazaar or open market than receivers served by various kinds of print or broadcast outlets. We need to re-think our ideas of message transmission, agenda setting, and framing. Using data from four decades of national (US) surveys, this paper will define and estimate the growth in just-in-time information acquisition in health and science. In recent decades, just-in-time information acquisition has grown to dominate health communication. Evidence suggests that adult science information acquisition is following a similar pattern, but still at an earlier stage.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Strategies for Improving Instruments and Increasing Data Quality

Making the most of your dry run with multi-use field testing: Examples from the 2017 Census of Agriculture field test

Jaki S McCarthy *USDA/Nat'l. Ag. Statistics Svc.*

Many large data collections will run a field test prior to operational field testing. This is typically used as a smaller scale dry run of the upcoming survey. The 2017 Census of Agriculture is the USDA National Agricultural Statistics Service's largest data collection, and is conducted once every 5 years. Prior to the 2017 COA, a field test is conducted as part of the multi method questionnaire evaluation. This large field test is used as a dry run of the questionnaire and data collection processes, but can also be used for other purposes. The 2015 field test was a crucial dry run, but was also used to produce records for the initial donor pool of records for imputation, compare the impact of alternative versions of the questionnaire on data quality and response rates, evaluate NASS's ability to accurately identify a pool of records for a reduced short form, test a new online system for web reporting, and test the ability of a pre-survey contact to help identify census nonrespondents. A field test is a rare opportunity that should be leveraged in multiple ways to make improvements to the subsequent data collections. Results from the 2015 COA field test and how they were used to benefit the COA will be discussed.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Enhancing the Quality of Internet Data Collection

How Long is Too Long? The Impact of Survey Length on Speeding, Cheating, and Falsifying in a Web Survey of Youth and Teens

Valrie M. Horton *Abt Associates*

Jared Knott *Abt Associates*

Benjamin Phillips *Abt Associates*

Rosaella Branson *Boys & Girls Clubs of America*

Elizabeth Fowlkes *Boys & Girls Clubs of America*

With the growing number of web surveys in the industry, survey length and the role it plays on data quality is of considerable interest, especially when surveying youth and teens. How long is too long? is the question we set out to explore in this research. In this paper, we examine survey length in a web survey context, using the Boys & Girls Clubs of America 2016 National Youth Outcomes Initiative (NYOI) member survey. The NYOI member survey is an annual survey of youth and teen members (ages 9-18) that measures academic success, good character and citizenship, healthy lifestyles, and club experience. With more than 145,000 online participants and more than 20 module combinations, the NYOI web dataset allows for detailed exploration of the impacts of survey length on data quality issues such as speeding, straight-lining and other patterned responses, and falsification. We utilize various respondent level data in this analysis. This data includes language (English or Spanish), age, grade, gender, race, device used for survey completion (computer or tablet), use of an audio component, and an honesty question. Data flags were created, post data collection, to identify fast respondents and respondents who straight-line or exhibit other patterned responses. We examine correlation between self-reported honesty, speed of survey completion, and patterned responses. In addition to the survey data, we perform exploratory research on cognitive development and attention span to give further advice when surveying youth and teens. Our conclusions shed light on maximum survey length for youth and teens in our research and provide guidance to other practitioners.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Issues in Opinion Formation and the Media

Assessing the Mood of the Nation Using Open Ended Questions: Challenges and Opportunities

Eric Plutzer *Penn State University*

Michael B. Berkman *Penn State University*

Burt Monroe *Penn State University*

Barak Obama's 2008 presidential campaign was built around a theme of hope. Identity based and nostalgia based social movements are built around feelings of pride. And, of course, feelings of anger or fear have played key roles in electoral contests. More broadly, emotions play a major role in attitude formation, attitude change, and attitude intensity; in trust and support for the political system and its specific institutions. As such, understanding how the public reacts emotionally to specific events – both political and non-political – is important. In this paper, we report on an innovative new poll that relies primarily on open-ended questions to tap respondents emotional responses to current events, and their hopes and fears for the future. The McCourtney Institute's / YouGov "Mood of the Nation Poll" is a periodic survey of 1000 Americans who each answer 4-5 questions that vary randomly in stem and in emotional prompt. Some ask about politics (e.g., What is there about American politics today that makes you feel proud?), some about the news (e.g., What has recently been in the news that makes you feel angry?), and some about the future, (e.g., Looking ahead, what worries you most about where America is headed in the next 12 months?). Each survey yields 5000 open-ended responses and over 50,000 words of text. Part 1 of the paper address challenges of data reduction – we compare hand coding based on a detailed topic rubric versus a machine learning topic model that derives clusters of themes based solely on the text and associated demographics. Part 2 of the paper reports on three key themes. We focus on how open ended responses enhance our understanding of (a) political polarization, (b) citizen distrust of government, and (c) how the Black Lives Matter movement is perceived differently by different citizens.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Data Collection in an International Context: Lessons Learned from the Field

When can we call?" Experiment to assess SMS text to prompt response across cultures

John Lee P Holmes *Qatar University Social and Economic Survey Research Institute (SESRI)*

Abdoulaye Diop *Qatar University Social and Economic Survey Research Institute (SESRI)*

Kien T Le *Qatar University Social and Economic Survey Research Institute (SESRI)*

Isam Abdelhameed *Qatar University Social and Economic Survey Research Institute (SESRI)*

Haneen B.K Alqassass *Qatar University Social and Economic Survey Research Institute (SESRI)*

Anis Miladi *Qatar University Social and Economic Survey Research Institute (SESRI)*

Abdulrahman Rahmany *Qatar University Social and Economic Survey Research Institute (SESRI)*

Yara Qutteina *Qatar University Social and Economic Survey Research Institute (SESRI)*

This paper examines a split sample experiment that alternates a Short Message Service (or SMS) advance text contact across seven language and three socioeconomic groups surveyed by wireless telephone in Qatar. Qatar's unique advantages include: a small size that facilitates population enumeration, universal cellular coverage, a multicultural respondent base, and a legal environment that permits such public contact techniques. Two years ago Qatar University's Social and Economic Survey Research Institute (SESRI) began conducting its own quarterly survey by phone to measure variation in quality of life and consumer confidence among its Qatari, high income expatriate ("white collar") and low income migrant ("blue collar") populations. Conducted using a random sample drawn from cellular telephones in Qatar, the survey has high response rates compared to surveys in the U.S. and Europe. Nevertheless concern about survey fatigue within a small population and the increasing use of avoidance features on smartphones has led SESRI to pre-emptively encourage positive response. To appreciate whether such a tactic would have the desired effect, advance text contact has been launched via a split sample experiment. Half of the sample receive an advance text sent out using Qatar University's SMS system and half do not. We measure whether this leads to increased blocking of SESRI's contact number, variance in contact and cooperation rates, and differences in the composition of the respondent pool by language group and respondent type (Qatari, white collar, blue collar). We believe this experiment has value in verifying whether such a method changes the composition of those who participate in the survey to better match the actual population. Because of the multiple language groups, it is applicable to broader cross-cultural survey quality concerns for other countries that may also seek to leverage widespread use of cellular and smart phone technology to understand popular preferences.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Interviewer Behaviors, Performance and Effects

Nurse effects in survey biomarkers

Joe Sakshaug *University of Manchester*

Alexandru Cernat *University of Manchester*

Surveys are extending the types of data they collect. An important new source is biological data such as blood or saliva collected from respondents. Nevertheless, this exciting new data source brings with it also a number of methodological issues. One of the most important ones is selection due to non-response and lack of consent to biological data collection. Our research aims to tackle this issue by looking at how nurses, respondents and areas influence 1) the likelihood that individuals will respond to the survey, 2) consent to biological data collection, and 3) the quality of the biological samples collected in two major British surveys: Understanding Society and English Longitudinal Study of Aging. This research will inform how nurses can have an impact on selection and data quality and how to model these effects as part of substantive research. It will also inform possible ways in which data collection could minimize nurse effects.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Data Collection in an International Context: Lessons Learned from the Field
#Censusfail 2016: Analysing the distribution of responses to the 2016 Australian Census, and what it tells us about privacy and other concerns
Nicholas Biddle *Australian National University*

In August 2016, the Australian Bureau of Statistics undertook the five-yearly Australian Census. This Census experience was unique in many ways. However, two main changes were that the default mode of collection was to be online, and that respondent names and addresses were to be retained for data linkage purposes for much longer than in previous years. These two changes resulted in considerable political and media discussion and a heightened attention to the Census. On the night the Census was supposed to take place, however, there was a series of server failures, leading to the online submission process being taken down that night and into the next day. Initial estimates suggest that, despite this, around 96 per cent of in scope households responded to the Census, slightly lower than the 2011 Australian Census, but higher than some other previous years. This headline figure does not, however, give any information on the distribution of item non-response, or the quality of information obtained. The aim of this paper is to use the Total Survey Error approach to analyse first release Census data, alongside geographically coded social media data (including from Google Trends) to gauge the effect of various decisions and incidents on the quality of Census-derived information. Inference will be drawn for collections of sensitive information by official statistical agencies, as well as analysis of such data by external researchers

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Data Collection in an International Context: Lessons Learned from the Field

Hanging up on F2F? Mode comparison for polling in Eastern Europe

Michelle E Romo *U.S. Department of State*

Marta Churella *U.S. Department of State*

Patrick Moynihan *U.S. Department of State*

Peyton Craighill *U.S. Department of State*

Face-to-face interviewing has been the main mode for high-quality survey research across Eastern Europe. As cell phone penetration increases across the continent, possibilities for moving to dual-frame telephone polling with high levels of coverage have improved. The Office of Opinion Research conducted multi-mode experiments in Poland and Ukraine to examine the feasibility of moving from face-to-face to telephone polling. Both the in-person and phone polls were designed to be nationally representative, used the same questionnaires, were conducted during the same time period, and had similar sample sizes within country (n=1,000 in Poland and n=1,500 in Ukraine). Different field firms administered each survey. This paper will explore mode differences and examine data quality variance across a variety of metrics including attitudinal and demographic differences, weighting efficiency, response dispositions, length of fieldwork, overall costs and general representativeness of the national populations. A key research question is to gauge the stability of trend data from in-person to phone, and whether stable political trends will be interrupted by a change in mode.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Interviewer Behaviors, Performance and Effects

Evaluating the Impact of CARI Rapid Feedback on Interviewer Performance on the American National Election Study (ANES)

Justin G Kamens *Westat*

Hanyu Sun *Westat*

Matthew DeBell *Stanford Univ.*

Vanessa Meldener-Harrell *Westat*

Survey management staff have long used Computer Audio-Recorded Interviewing (CARI) to monitor field interviewers' work, whether for assessing interviewer performance, validating interviews, or evaluating survey questions. CARI allows studies to monitor interviewer performance in real time during the field data collection period. In recent years, Westat has developed a system and set of procedures designed to provide feedback to interviewers on problematic interviewing behaviors soon after they are detected. In this Rapid Feedback program, experienced quality control staff listen to the CARI audio recordings of field interviews to code interviewers' performance and behavior on a variety of criteria, including reading questions verbatim to respondents, correct probing behavior, and recording responses that match the respondent's answer. Interviewers that display negative behaviors then participate in a personal feedback session with a quality control staff member, with the goal of correcting these problematic behaviors before they are continued into future interviews. This program effectively extends interviewer training into the data collection period and facilitates corrections to interviewer behavior that may affect data quality in future interviews. These procedures were implemented during the Pre-election data collection phase of the 2016 American National Election Study (ANES), with cases coded for 107 field interviewers. Every interviewer's first completed case was coded via CARI. Feedback sessions were scheduled for interviewers with poor performance, and then a second case was coded for those interviewers to examine the effectiveness of the Rapid Feedback. This paper will evaluate the impact of the Rapid Feedback program on interviewer performance on ANES. We will examine the interviewer's behavior changes between the first and second coding, the presence of positive and negative interviewing behaviors, as well as quality control staff's evaluation of the two codings. Implications for future improvements to CARI feedback procedures will be discussed.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Interviewer Behaviors, Performance and Effects

One size does not fit all: Measuring and evaluating interviewer performance in multiple dimensions

Hanyu Sun *Westat*

Gonzalo Rivero *Westat*

Matthew DeBell *Stanford Univ.*

How to measure the performance of workers is a common problem that arises in all industries. The survey field is no different. The common consensus among survey practitioners is that the skills and characteristics of the field interviewers have a strong and direct impact on the quantity and quality of the data collected. However, very little research has attempted to systematize the measurement of interviewer performance or to offer the guidance on what attributes to look for when select and evaluate field interviewers. We attempt to fill this gap with the current study. We draw from the literature in labor econometrics to identify and empirically measure two (possibly uncorrelated) dimensions of interviewer performance: the ability of the interviewer to gain cooperation from the respondent to complete an interview, and the ability of the interviewer to collect survey data in good quality. We will use data from the face-to-face portion of the 2016 American National Election Studies (ANES), including interviewers' demographics, paradata, performance scores of a sample of recorded interviews, evaluations of the interviewers' training performance, and an assessment of the quality of the survey data. We then will identify each dimension separately to estimate the empirical trade-off between the two. Finally, we will compare our individual estimates to the rankings produced by the performance metrics most commonly used in practice (i.e., the ratio between the number of completed cases and the hours worked for a particular interviewer). This research will have implications for the recruitment of field interviewers and for the development of the interviewer performance metrics to be used during the field period.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Issues in Opinion Formation and the Media

How Americans Navigate the Modern Information Environment

Jennifer Benz *NORC at the University of Chicago*

Norman Bradburn *NORC at the Univ of Chicago*

NORC at the University of Chicago conducted a national study that provides a detailed understanding of Americans' perspectives on an information environment where most have instant access to virtually unlimited amounts of information via the internet. This study addresses questions about how people actually use these sources and how they react to the steady streams of sometimes conflicting information they confront daily. The study reveals that while the vast majority of Americans believe it is easier to find useful information today than it was five years ago, 78 percent report the sheer quantity of information can sometimes be overwhelming. Legacy media continue to fill important roles for Americans. For example, while Americans are most likely to go online for information when making a major product purchase, they are more likely to turn to TV when they need information to decide where they stand on a major national issue. Further, Americans across demographic characteristics report a healthy amount of skepticism over the trustworthiness of online information, especially when found through social media. Taking a deeper look at Americans' information habits, whether deciding on a major purchase or their stance on a policy issue, many Americans rely on gut feelings to determine when information is suspect. How Americans view source credibility and expertise about national issue information has important implications for their policy attitudes. Even controlling for a person's political affiliation and other demographic factors, those Americans who place a higher value on sources that refer to experts, scientific evidence, government data, or other credible sources hold more supportive policy preferences on issues such as legalization of same-sex marriage, belief in global warming, and support for the ACA. The nationally representative survey of 1,007 adults was conducted from 1/14-31/2016 using the AmeriSpeak Panel®.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Enhancing the Quality of Internet Data Collection

Tailoring Nonresponse Follow-up in a Web Survey of College Students Based on Predicted Propensity to Respond

Rebecca Medway *American Institutes for Research*

Mengmeng Zhang *American Institutes for Research*

Mark Masterton *American Institutes for Research*

Evan Nielsen *American Institutes for Research*

As survey response rates decline, conducting high-response surveys has become increasingly challenging and expensive. Simply increasing the number of survey contacts in an attempt to increase response rates may do little to reduce nonresponse bias – at worst, it may increase bias if these contacts simply bring in more of the same kinds of respondents. These issues have prompted increased consideration of tailored designs – survey designs that aim to use resources efficiently by targeting follow-up protocols to specific subgroups based on their predicted likelihood of responding to the survey request. This presentation will report on the results of a tailored nonresponse follow-up experiment that was included in a recent web survey of college students. Propensity to respond was estimated for all students using a regression model that predicted response to a prior administration using variables available on the frame. The students were grouped into four propensity groups (high, medium, low, very low), and half the students in each group were randomly assigned to either a standard or tailored nonresponse follow-up condition. In the standard condition, all students received the same nonresponse follow-up contacts. In the tailored nonresponse follow-up condition, students received varied contacts based on their predicted propensity to respond, with more resource-intensive contact methods, such as telephone, being targeted at lower propensity students and being avoided for higher propensity students. Preliminary results suggest that the response rate was equivalent in the two conditions; however, there was increased representation of lower propensity students in the tailored nonresponse condition. We will discuss the impact of the tailored contacts on outcomes such as respondent characteristics, survey responses, and response quality (e.g., item nonresponse, nondifferentiation). We also will discuss success, challenges, and lessons learned from attempting to use telephone and text to contact college students about completing an online survey.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Issues in Opinion Formation and the Media

Searching for News: The Flint Water Crisis

Katerina Eva *Matsa Pew Research Center*
Andrew Mercer *Pew Research Center*
Galen Stocking *Pew Research Center*

Americans' search habits can reveal a great deal about the news that different segments of our population want to learn more about, how they make sense of the news of the day, and what concerns may be most pressing. This proactive news-seeking behavior represents a different kind of engagement with the news from passive exposure on a social media feed or in the newspaper. We explore this phenomenon through a case study of the water crisis in Flint, Michigan between 2014 and 2016, when a change in Flint's source of tap water threatened public health through contamination and lead poisoning. The central research question is: How has the water crisis in Flint, MI impacted the public's interest in learning more about what happened and about water quality more broadly? Using data from the Google Health API, we evaluate how the public changed its internet search patterns in response to media coverage of the Flint water crisis over the course of two years. Our analysis groups searches into the following categories: personal health, the environment and public health, contaminants, political developments and news updates. The Google Health API allows us to compare search patterns over time and across different areas of the country in ways that are not possible with the publicly available Google Trends API. In particular, we compare search patterns in Flint and its surrounding areas to patterns observed nationally and in Michigan. We discuss our substantive findings as well as some of the methodological challenges we encountered while working with search data from the Google Health API.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Issues in Opinion Formation and the Media

PANJAAPOR Student Paper - Amplifying the Effects of Winning and Losing: Partisan Media's Effects on Perceived Electoral Integrity in the 2008 and 2012 Presidential Elections

Andrew M Daniller *University of Pennsylvania*

Citizens who voted for the losing side in an election are much less likely to believe that the process was fair than citizens who voted for the winner. However, little attention has been paid to partisan media's potential to exacerbate this phenomenon. Partisan media are likely to condition the effects of winning and losing through two mechanisms: expectations and explanations. I hypothesize that like-minded media amplify the effects of winning and losing on perceptions of electoral integrity. In other words, supporters of a winning candidate or party become even more trusting of the process when exposed to media that favors their side, while supporters of a losing candidate or party become even less trusting of the process when exposed to media that favors their side. At the same time, I hypothesize that cross-cutting media mute the effects of winning and losing, decreasing the magnitude of changes in perceptions of electoral integrity. I test my predictions using nationally representative panel survey from the 2008 and 2012 presidential elections, finding that like-minded media amplified the effects of losing but not of winning in these election cycles. Additionally, random assignment of post-election interview dates allows me to test whether the proposed mechanisms are consistent with the observed effects of partisan media. The data suggest that partisan media influence perceptions of electoral integrity by influencing citizens' pre-election expectations, but not necessarily their post-election explanations for the outcome.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Enhancing the Quality of Internet Data Collection

The Effect of Respondent Commitment and Tailored Feedback on Response Quality in an Online Survey

Kristen Cibelli Hibben *Univ of Michigan/Program in Survey Methodology*
Frederick Conrad *Univ. of Michigan, ISR*

To the extent that the inaccuracy in survey responses is due to insufficient effort by respondents, it might help to obtain a direct commitment from the respondent to be diligent in completing the task and to provide feedback “tailored” to the respondent’s effort on the response process. The rationale for respondent commitment is that agreeing or stating one’s intention to behave in a certain way commits a person to carry out the terms of the agreement. Tailored feedback informs respondents about the expected response process (e.g. “You answered that a bit quickly”). Education theory and existing research suggests that feedback can curtail negative behavior (e.g. speeding). Charles Cannell and his associates pioneered these techniques in the late 1970s in face-to-face health-related interviews and the results were promising: The quality of reporting on response tasks increased with asking respondents to commit to providing complete and accurate information and the use of interviewer feedback (Miller & Cannell, 1977, 1982; Oksenberg, Vinokur, & Cannell, 1977a; Oksenberg et al., 1977b; Vinokur, Oksenberg, & Cannell, 1977). The current study examines the effects of respondent commitment and tailored feedback on the accuracy of reported medical visits in an online survey of the parents of child patients at University of Michigan Health System carried out in March-May 2016. Response accuracy to several health service utilization questions is validated using medical record data – previous studies evaluating commitment and tailored feedback have only used indirect measures of accuracy. The study also examines the effect of respondent commitment and tailored feedback on item nonresponse, socially desirable reporting, satisficing, respondent-reported record check, and debriefing questions. Applying these techniques to an online survey seizes an important opportunity to draw on the interactive features of the web and may offer a cost-effective means of improving web survey quality.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

90210 Isn't Hollywood! Or Is It? Issues Related to Geographic Specificity of Cell Phone Samples

Zip Code and Cellphone Area Code Mismatch: Implications for Health and Survey Research

Beatrice Abiero *Ipsos Public Affairs*

Robert Petrin *Ipsos Public Affairs*

Landlines have historically played a major role in health and survey research. However, the research industry must continually account for the implications of ever-increasing cellphone-only ownership over landline use when designing health and other survey research. One of the key considerations with respect to contacting cellphone-only users is the inaccuracies (or mismatch) in geographic information between zip- and area codes, making it difficult to “place” cell phone numbers, even when specially adjusted cell phone frames are used. Given that many health outcomes are spatially-situated, this geographic mismatch has important implications for sample design and subsequent data analysis. For example, recent research shows that area- and zip code inaccuracies are more prevalent among highly mobile demographics such as college students, seasonal and temporary workers, and renters. More mobile populations are in turn exposed to different sets of health risks, behaviors, and constraints than less-mobile populations. Additionally, findings by the National Center for Health Statistics suggest that individuals in cellphone-only households tend to report more sedentary and risky health behaviors (e.g., smoking, binge drinking) compared to individuals in landline-only or cellphone and landline households (Blumberg & Luke, 2016). This paper first identifies the extent of zip- and area code mismatch based on data drawn from a large national proprietary panel survey. Federal data was used to contextualize findings from this study in order to relate this mismatch to challenges in evaluating individual health outcomes, as well as differences in regional community health profiles (particularly for pockets with high zip- and area code mismatch). Finally, this study proposes techniques relevant for correcting zip- and area code mismatch for sampling design and analysis in health and survey research.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Strategies for Improving Instruments and Increasing Data Quality

Validating and Assessing Drug Use Self-Reports In a Longitudinal Youth Study

Vanessa E Thornburg *RTI International*

Diana Fishbein *The Pennsylvania State University*

Ty Ridenour *RTI International*

Lilia Fillepenko *RTI International*

Jaki Brown *RTI International*

Accurate measures of drug use in study populations are essential to research studies that focus on their effects. Study staff must have confidence in the validity of their study population's drug use self-reports. To this purpose, the prospective longitudinal Healthy Adolescent Development Project Sleep Study (HAD Sleep Study), which examines neurocognitive susceptibility factors and consequences of drug abuse among 529 adolescents in a small community with high prevalence rates of early adolescent marijuana use, embedded a self-report validation sub study within its current third wave of data collection. The validation sub study protocol calls for youth to complete a urine metabolite 5 panel drug test that aligns with those youth drug use self-report (ACASI). The 5 drugs in the urine panel are those most commonly called out as associated with negative neurocognitive outcomes. In addition to providing an opportunity to test the validity of youth self-reports, this validation sub study examines factors that could affect HAD Sleep Study youths' willingness to accurately report their drug use, including youth risk profiles, and gender and race effects, and social pressures and interviewer effects influencing rates of self-reporting. We review the results of our validation analyses and offer insight into the accuracy of self-reports and factors that affect accuracy, to assist in designing research studies that rely on drug use self-reports.

May 19th, 2017
8:00 AM - 9:30 AM
Concurrent Session B

Issues in Opinion Formation and the Media

Expectancy Violation and Costly Signaling: An Interactive Framework for Political Persuasion

Victoria Dounoucos *Duke University*

Research into political persuasion has primarily focused on either source or message factors, with studies attempting to identify factors in evaluating the persuasive effects of political messages. This research has shown that a speaker's credibility – measured by trustworthiness and expertise - is highly important in determining his or her persuasiveness. This paper expands upon the extant literature by considering the interactive effects of message and source factors. I argue that the interaction of source and message effects, particularly in situations in which these combined effects violate expectations, is what makes political displays of unexpected support successful. I examine this interaction by addressing two research questions: (1) how do source and message cues affect an audience's evaluations of a speaker; and (2) what effect do these cues have on issue opinion formation. I find that when individuals are faced with an unexpected message, they evaluate the source as more trustworthy and credible than the source delivering an expected message. Importantly, this unexpected message results in a shift in issue attitudes in the direction of the speaker's promoted stance.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

AAPOR Election Review Panel

Panel for the AAPOR Election Review Group

Courtney Kennedy *Pew Research Center*

Kristen Olson *University of Nebraska*

Evans Witt *PSRAI*

Doug Rivers *Stanford University/YouGov*

Kyley McGeeney *Penn Schoen Berland*

Claire Durand *University of Montreal*

Joshua Clinton *Vanderbilt University*

Scott Clement *Washington Post*

Mark Blumenthal *SurveyMonkey*

Lydia Saad *Gallup*

Charles Franklin *Marquette University*

Election years present particularly high profile moments for survey research. This is a time when polls dominate the media and the accuracy of polls can be confirmed or refuted by the actual vote outcome. Frequently, this comes with it a chorus of concerns about “a crisis in polling.” And polling mistakes make better headlines than do on-target polls. Understanding and being able to articulate the overall outcomes of election polling, the changing methodologies being used, and the potential for variation in the accuracy of polls is vital for our industry. To support the collection and dissemination of such information, AAPOR convened a panel of survey research and election polling experts to conduct a post-hoc analysis of the 2016 pre-election polls. The goal of this committee was to prepare a report that summarizes the accuracy of 2016 pre-election polling (for both primaries and the general election), reviews variation by different methodologies, and identifies differences from prior election years. The report is intended to help the industry, the media, and other interested parties to have a clearer picture of the successes and the misses, and what likely caused each, during the 2016 election polling cycle and provide groundwork for discussion of polling in the 2020 election cycle. This panel presents the main findings from the report and will facilitate a discussion about the implications for AAPOR and the polling industry more generally.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Methodological Brief: Writing Questions to Increase Data Quality

Panel Conditioning in Measuring Ego-Centered Social Networks in Online Surveys

Henning Silber *GESIS - Leibniz-Institute for the Social Sciences*

Jette Schröder *GESIS - Leibniz-Institute for the Social Sciences*

Bella Struminskaya *Utrecht University*

Michael Bosnjak *GESIS Leibniz Institute for the Social Sciences*

Social scientists are increasingly interested in social networks of respondents, and many large-scale surveys include a network module, such as the GSS and the ANES study. Questions for collecting network data usually start with name generators: respondents name the persons to which the question applies. In a second step, additional information about the relationship of the persons in the network is collected. As a result, the number of questions a respondent is asked increases with every person added. One major threat to the validity of the findings collected with such social network modules is the tendency of some respondents to reduce interview time by underreporting the number of contacts. This threat is especially connected to longitudinal surveys encompassing repeated network modules used several times. Changes in respondents reporting behavior as a consequence of the previous survey participation is referred to as a panel conditioning effect. The overall aim of this study is to investigate if, and to what extent, panel conditioning effects reduce the size of self-reported ego-centered networks. Moreover, we seek to identify moderators of the effect, such as respondents' motivation and ability. We conducted an experiment with a non-probability online panel of German adults ($n = 552$) in 2016. Respondents were randomly assigned to two experimental groups. In the first wave, only the experimental group received the social network questions. In the second wave, experimental and control group both received the social network questions. This experimental between subject design allows us to test the hypotheses whether the number of contacts and number of relationships within the network decreases due to panel conditioning. We will report different indicators of data quality such as the number of contacts, the number of relationships within the social network, item non-response, straightlining, and response times.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Methodological Brief: Writing Questions to Increase Data Quality
Continued Evidence on Clarifying Instructions Improving Response Rate and Quality of Numerical
Open-ended Questions

Anil Kumar Chaudhary *University of Florida*
Glenn Israel *Univ. of Florida*

Compared to face-to-face and telephone interviews where interviewer assist respondents, in self-administered surveys, respondents make meaning of questions on their own. Sound visual design is always needed to assist respondents to efficiently complete self-administered surveys. Verbal instructions are tested to improve item response rate when they are provided with sound visual design. Visual design's effect on item response rate and measurement error is well-tested but evidence of effect of clarifying instructions on item response rate and response quality is limited. We designed an experiment in 2015 and replicated the same experiment in 2016 with slight modification to understand how clarifying instructions (with/without) affect response to two numerical open-ended questions. Data for years 2015 and 2016 were collected from a web/mail mixed-mode annual customer satisfaction survey of Florida Cooperative Extension Service (FCES). Two numerical open-ended questions asked respondents information about number of times they contacted FCES in past 12 months and years using Extension services. The clarifying instructions used with two questions in 2015 were "include 1 for the time that this survey asked about" and "write '1' if only this year" respectively. While in 2016, the instruction for years using Extension services was same, but for contact question we revised instruction as "include 1 for the time when you received information : piped in information." Over the period of two years we found consistent finding that addition of clarifying instructions significantly reduced the percent of missing and incorrectly formatted responses for both questions. For response mode, main effect of response mode (paper, web response with laptop and desktop, and web response with mobile devices) was significant in year 2015, while for year 2016 main effect of response mode was marginally significant.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Methodological Brief: Writing Questions to Increase Data Quality

Can We Improve the Way We Ask about Health Insurance Coverage and Health Care Spending?

Findings from an Experiment

Susan A. Sherr *SSRS*

Sharon Long *Urban Institute*

Zi Zhang *CHIA*

Huong T. Trieu *CHIA*

Kathy Langdale *SSRS*

Surveys about health care and health insurance coverage routinely ask respondents to quantify their health care spending and outstanding health care debt. This is generally motivated by a desire to calculate the burden that spending on health care over and above what might be paid for by health insurance places on consumers and to determine the prevalence of underinsurance, where the respondent is insured but lacks sufficient financial protection from high health care costs. Although respondents are asked to report spending and often will provide estimates of their expenditures and debts, it is an empirical question as to the accuracy of these estimates and the degree of confidence respondents actually have in their reported numbers. In 2016, respondents to the 2015 Massachusetts Health Insurance Survey, a general population survey of the Massachusetts population administered by the Commonwealth since 2008, who agreed to be recontacted were interviewed about the financial elements of their health insurance coverage (e.g., deductibles), out of pocket health care spending, and unpaid health care bills. Respondents were randomly assigned to be either asked questions about amounts of deductibles, health care spending and debt in an open-ended format or to be provided with ranges of dollar amounts from which to select their responses. All respondents were then asked to report on their level of confidence in the information they provided. Analysis includes comparing the degree of non-response between the two groups and determining whether there are statistically significant differences in confidence levels. This methodology brief will describe findings from the experiment and report on how results can inform the design of future health coverage instruments to improve the accuracy of reporting on the financial aspects of health insurance coverage and health care spending.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Methodological Brief: Writing Questions to Increase Data Quality

Casting a wide net: Specification error in screening homeschool children

Danielle Battle *American Institutes for Research*

Mahi Megra *American Institutes for Research*

Specification error, a type of non-sampling error, occurs when the concept intended to be measured is different from what is actually measured. For example, a screening question asking “how many adults live in this household” could have different interpretations. Respondents may count students away from home as living in the household when they should not count them. Unless the respondent can clearly interpret screening questions, the responses may bring in unintended participants for a follow-up survey. In the 2016 National Household Education Surveys (NHES:2016), a two-phase self-administered mail and web-based survey, the screener respondent enumerated the members of the household and indicated the school enrollment status of each member. Based on the response to the school enrollment question, households with school-aged children were sampled for one of two topical surveys: one for homeschooled students or one for students enrolled in public or private schools. In NHES:2016, due to concerns that prior administrations may have missed part of the homeschooled population, the order of the response options for the screening item was changed so that the “homeschool” option was presented before the “public or private school” option, possibly making it less likely for respondents to overlook. However, preliminary findings from the NHES:2016 administration indicate that, unlike in prior administrations, the response rate was lower and breakoff rates were higher for the homeschool survey than for the enrolled-students survey. This suggests that respondents may have misunderstood the homeschool response option when it was placed before the “public or private school” option; thus, paradoxically, NHES:2016 may have sampled more households that do not have homeschooled children. By reviewing responses to topical survey items related to the definition of a homeschool child, this analysis will examine whether specification error in the screening question resulted in households being incorrectly sampled for the homeschool survey.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Methodological Brief: Writing Questions to Increase Data Quality

A comparison of ranking order methodologies

Angela Fontes *NORC at the Univ of Chicago*

Angela Akinyemi *JUST Capital*

Rob Brown *JUST Capital*

Can Geng *NORC at the University of Chicago*

Rene Bautista-Martinez *NORC at the Univ of Chicago*

Kristin Dwan *NORC at the University of Chicago*

The ordering of multiple items is an ongoing challenge in public opinion research. There are several methods used to develop a ranked list of items. Traditional methods include simple ranking of all items, and points allocation exercises, where respondents are presented all items and asked to assign points indicated the importance of each item. However, previous research indicates the ability of respondents to perform these type of tasks, particularly with long lists or complex items, is limited. As a consequence, researchers have developed alternative methods of developing ranked lists. In this study, we explore four methodologies, using data from three surveys, to examine the effectiveness of ranking approaches. As part of a series of surveys investigating public perception of corporations and businesses, we measure how Americans understand ethical corporate behavior (what we call 'JUSTness'), and how these perceptions impact decisions on purchasing, investing, and employment. We compare the results of three surveys within this series, utilizing two techniques used in developed ranked lists. All surveys examined the rank order of the same ten categories of JUSTness. Two surveys were collected using a probability-based panel, and the third using an opt-in panel. The first survey asked respondents to rank four items out of ten in a single question. Data from the respondent assigned ranking was then used to estimate utility weights using a maxdiff approach. The second study developed the same maxdiff estimates with a fully balanced maxdiff design (where all possible combinations of items are presented, and the number of times each respondent is presented each item is consistent). Results from these four approaches will be compared and discussed in the paper.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Methodological Brief: Writing Questions to Increase Data Quality

The KISS Principle in Survey Measurement: Results from the General Social Survey

Duane Alwin *Pennsylvania State Univ.*

Brett Beattie *Match.com*

Erin Baumgartner *Rice University*

The KISS principle emphasizes simplicity of design. KISS is an acronym for “Keep it simple, stupid” stressing the idea that most things work best if they are kept simple rather than made complicated. Variations on the phrase include “Keep it short and simple,” “Keep it simple and straightforward,” or “Keep it simple, silly,” all of which emphasize the goal that unnecessary redundancy and complexity should be avoided and the achievement of perfection depends on parsimony. This idea is revered in science, where Occam’s razor dictates that the simplest explanation is preferred over more complex ones. In survey design, the KISS principle may be applied to the design of questions and questionnaires—specifically the length of questions, the inclusion of questions in batteries rather than stand-alone questions, and the number of questions in batteries. To this point, while there is a variety of expert opinion on such matters, too little empirical information is available on the relationship between the simplicity of design and the quality of survey data. To fill this gap, our research tests several hypotheses about aspects of survey design using reliability of measurement as the criterion for evaluating measuring precision. Using data from three General Social Survey (GSS) panel studies—2006-08-10, 2008-10-12, and 2010-12-14—we assess data for over 500 survey questions with respect to the reliability of measurement. Consistent with our expectations, the preliminary findings suggest that, at least with respect to some types of survey questions, there are declining levels of reliability of measurement for more lengthy questions. Along the same lines, we also find that questions in batteries are somewhat less reliable, and even more so for lengthy batteries of questions. We conclude that survey design should pay heed to the principle of simplicity, and we suggest that survey researchers should “KISS”.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Methodological Brief: Writing Questions to Increase Data Quality

Is the Net Promoter Score (NPS) an appropriate metric to assess satisfaction for internal enterprise applications? Validity of and methodological considerations for NPS usage in an enterprise setting.

Randal R Ries *IBM*

Annette Tassone *IBM*

Felix Portnoy *IBM*

Danny Hager *IBM*

Claude Elie *IBM*

James Newswanger *IBM*

Mark Wise *IBM*

The Net Promoter Score (NPS), introduced by Reichheld in his seminal article “The One Number You Need to Grow” (HBR, 2003), requires participants to answer the question “How likely are you to recommend this [product/company] to a friend or colleague?” on an 11-point scale (from “0 – Not at all likely” to “10 – Extremely likely”). This metric has been widely adopted in enterprise and consumer industries for its ease of use and power to assess customer loyalty. “Overall satisfaction” (the percentage of respondents who indicate they are “somewhat satisfied” or “very satisfied” on a 5-point Likert scale), has also been commonly utilized to assess sentiment. In an analysis conducted within a large global enterprise where both measures were collected, high user satisfaction was found in tandem with negative NPS scores – raising concerns about their validity. In this paper, we address these concerns and provide evidence that NPS can be an effective metric for assessing user sentiment even for internal enterprise applications (where there typically is not a product choice). Additionally, we address methodological issues (e.g., 5 vs. 11 point scales, question order when both are assessed, etc.) and provide an analysis of the relationship between satisfaction and NPS.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Increasing Data Quality in Web Surveys: Predicting and Managing Undesirable Respondent Behaviors

Motivated Misreporting in Web Panels

Ruben Bach *Institute for Employment Research*

Stephanie Eckman *RTI International*

Previous studies of reporting to filter questions have shown that different filter question formats lead to different responses. While respondents in a grouped format are asked follow-up questions only after multiple filters have been administered, respondents in an interleaved format are asked follow-up questions immediately after each filter. The common interpretation of this difference in reporting is that respondents surveyed in the interleaved format learn to say "No" to the filter questions in order to shorten the interview. However, so far, no study has examined whether misreporting to filter questions worsens over time in a panel study. We conducted an experiment using filter questions in two consecutive waves of a monthly online panel, which allows us to study how misreporting to filter questions changes over time. While we replicate previous findings on the format effect, we find do not find any support for the hypothesis that respondents' reporting behavior gets worse over time. Our findings add to the literature on data quality in web panels, panel conditioning and motivated misreporting.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Obtaining Permission and Consent for Hard to Obtain Health Data

Questionnaire Experiments to Increase Provider Consent

Vince Welch *NORC at the University of Chicago*

Megha Ravanam *NORC at the University of Chicago*

Benjamin Skalland *NORC at the University of Chicago*

Renee Ramsey *NORC at the University of Chicago*

Sarah Reagan-Steiner *Centers for Disease Control and Prevention*

Cynthia Knighton *Centers for Disease Control and Prevention*

Holly Hill *Centers for Disease Control and Prevention*

Laurie Elam-Evans *Centers for Disease Control and Prevention*

Chalanda Smith *Centers for Disease Control and Prevention*

The National Immunization Surveys are dual frame random-digit-dial telephone surveys used to monitor vaccination coverage among children age 19-35 months and teens 13-17 years of age in the United States, based on provider-reported vaccinations. The surveys collect household-reported demographic and access to care data during telephone interviews with the parent or guardian and then ask for consent to contact the child's vaccination provider(s) to obtain provider-reported immunization history using a mailed questionnaire. Consent to contact the child's vaccination provider(s) is crucial for the success of the surveys, which base estimates on provider-reported vaccinations, except for influenza vaccination of children 6 months-17 years. Consent has been decreasing over the last several years; therefore, it is important to explore methods for improving provider consent. Work by Tourangeau and Ye (2009) and Sakshaug (2014) has shown that the use of "loss-framing" language can have a positive impact on consent rates. Loss-framing language emphasizes that the first data collection effort will be less useful without the second. Two studies were conducted to determine the effect of three independent experimental treatments in web modes and in computer-assisted telephone interviewing (CATI) on provider consent rate. The experiments were both full factorial 2 x 2 x 2 experimental designs. The web and CATI experiments both tested the use of loss-framing language and the inclusion of normalizing language (which emphasizes that most respondents provide consent). The web experiment tested the placement of the request for consent while the CATI experiment tested early notification that this would be a two-part survey. Although other studies have shown an increase in consent rates with the use of loss-framing language, the two experiments did not see any effect on provider consent rates. We will discuss our findings and also examine and discuss possible mode effects.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Leveraging Contact Strategies and Response Propensities to Increase Survey Participation

Developing Estimates of Sampling Variability for the Planning Database's Low Response Score

Luke J Larsen *U.S. Census Bureau - Washington, DC*

The Planning Database (PDB) is an annual Census Bureau data product that contains population and housing statistics about the United States at the tract and block-group levels of geography. The contents of the PDB come from the publicly available Summary Files of the most recent decennial census and the most recent American Community Survey (ACS) 5-year file. Though the PDB has many applications, one of its chief uses is field operations planning for various Census Bureau survey projects, including the decennial census. To that end, the Census Bureau includes on the PDB the Low Response Score (LRS), which is an estimate of 2010 mail return rates by geography (Erdman and Bates, 2014). The LRS is based on an ordinary least squares (OLS) regression model that is fit using PDB data. Using the LRS, survey planners can identify which geographies may require more effort to enumerate and can allocate resources accordingly. Because the LRS is calculated using ACS data, it will have sampling variability; however, the LRS is currently produced without associated measures of error. This is important because the PDB is published at sub-county levels of geography, where ACS-based estimates often have increased sampling variability. Therefore, having both the LRS and associated variance estimates are crucial for data users to make informed decisions during the survey planning process. In this paper, the author uses ACS 5-year microdata to recreate the original OLS model while accounting for sampling error, and then uses that model to generate LRS estimates with margins of error. From these results, it is possible to study coefficients of variation for the LRS across the entire United States and gain a sense of volatility for this important metric.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

B.Y.O.P.: Build Your Own Panel

Experiments in Recruiting the Life in Australia Probability-based Online Panel

Graham M Challice *Social Res Ctr Pty Ltd*

Paul Lavrakas *Independent Consultant*

Lars Kaczmirek *GESIS - Leibniz Institute for the Social Sciences*

Darren W Pennay *Social Research Centre Pty Ltd*

Probability-based online panels emerged in the USA in 1999 with the Knowledge Networks Knowledge Panel and in Europe in 2007 with the Longitudinal Internet Studies for the Social Sciences (LISS) panel in the Netherlands. In Australia, despite high internet penetration and the rapid take-up of online research by the market and social research industry, the first national probability-based online panel didn't appear until 2016. The desirability of Australia having a probability-based online panel was demonstrated by the Australian Online Panels Benchmarking Study (OPBS); see Pennay et al., 2016. The OPBS comprised eight national surveys – three probability surveys and five surveys conducted via non-probability online panels. Findings replicated earlier international research (e.g. Yeager et al., 2011) in showing that non-probability sample surveys done via the Internet were consistently less accurate, on average, than probability sample surveys. This research led to the creation of the probability-based Life in Australia online panel, which is being established using dual-frame RDD telephone surveying. This paper reports on the results of the recruitment trial which controlled for two conditions using a 2 x 2 experimental design: 1) The placement of the recruitment question was alternated to before (ask first) or after (ask last) a series of profiling questions and; 2) the nature of the recruitment request alternated between attempting to recruit panel members directly over the phone (direct approach) or asking respondents to accept a recruitment pack via mail or email with a view to then joining the panel (indirect approach). The effectiveness of these approaches and that of the final recruitment method used to build the panel are described in this paper. The results from a replication of the OPBS undertaken on the Life in Australia panel is also reported.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Leveraging Contact Strategies and Response Propensities to Increase Survey Participation

Modeling Contact and Cooperation in an Administrative Record Based Mail/CATI Survey

Frost Hubbard *IMPAQ International*

Cheryl Wiese *IMPAQ International*

Amy Djangali *IMPAQ International*

Surveys whose sampling frames come from administrative records contain a rich set of sociodemographic data, making them fertile ground for nonresponse bias analyses. In addition, sequential, mixed-mode surveys based on a frame of administrative records allow researchers to explore how the introduction of different modes affects nonresponse. We use the 2016 Kidney Disease Quality of Life Survey (KDQOL), whose sampling frame consisted of administrative records of Medicare beneficiaries of End Stage Renal Disease care receiving treatment in specific areas of the country. We construct logistic models to analyze effects of sociodemography on the different components of nonresponse, non-contact and non-cooperation, in each data collection phase of the sequential, mixed-mode design. The survey first mailed questionnaires to the best known address of all beneficiaries and then conducted centralized telephone follow up calls with all cases that did not respond to the mail survey request. Some sociodemographic groups were less likely to have deliverable mailing addresses and to respond by mail if we could deliver the survey. Among the beneficiaries that did receive at least one phone call, some sociodemographic groups were less likely to have a working telephone number, conditional on the beneficiary having a deliverable mailing address. Based on the findings, we consider data collection strategies to help reduce both non-contact and non-cooperation within each mode for this specific population.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Measurement Issues Surrounding the Provision of Health Care and the Affordable Care Act (ACA)

Insure My Beating Heart: Exchange Participant Characteristics and Satisfaction With the System

Stephanie Marken *Gallup*

Zachary Auter *Gallup*

In 2008, Gallup and Healthways initiated a partnership to track and understand the key factors that drive well-being. The Gallup-Healthways Well-Being Index provides an in-depth, nearly real-time view of Americans' well-being. The Gallup-Healthways Well-Being Index is a nationally representative survey of adults currently living in the United States interviewed by telephone every day, 350 days per year. Since this study was initiated in 2008, Gallup has measured the insurance status of adults currently living in the United States as part of the Gallup-Healthways Well-Being Index survey. Gallup's longstanding trend on this important metric provides the ability to analyze the insurance rate prior to during and now following the implementation of the Affordable Care Act (ACA). In addition to tracking the insurance status of those currently living in the United States, Gallup closely monitors the percent of newly insured Americans that obtained their insurance through the exchanges, and public reaction to and support of the ACA including the perceived impact of the policy on Americans and their families. This presentation will provide researchers important insights into how the uninsured rate has changed over time and how satisfied or dissatisfied those who obtained their insurance through the exchanges are with the healthcare system. We will also provide details about the demographic characteristics of those who obtained their insurance through the exchanges, compared with those who obtained their insurance through other sources including their self-reported health status (excellent, good, fair or poor), age, gender and race/ethnicity.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Reducing Barriers Due to Language and Coverage

Reaching out to Spanish-speaking Respondents in an IRS Household Survey

Jocelyn Newsome *Westat*

Jennifer Anderson McNulty *Westat*

Kerry Levin *Westat*

Brenda Schafer *Internal Revenue Service*

Patrick Langetieg *Internal Revenue Service*

Saurabh Datta *Internal Revenue Service*

In recent years, survey methodologists have sought to increase response from Spanish-speaking respondents. About 16 million people in the United States are Spanish speakers with no or very limited English proficiency, and studies have shown that mail surveys are likely to underrepresent Spanish speakers (Caporaso et al., 2013)—particularly when materials are presented only in English. In one study, response rates among Spanish speakers were half that of English speakers (McGovern, 2004). While sending survey materials in both English and Spanish to all respondents has been shown to increase Spanish response (Brick et al., 2012), this approach can be prohibitively expensive. The IRS Individual Taxpayer Burden (ITB) survey is an annual survey sent to 20,000 individuals in the United States. It measures the time and money taxpayers spend complying with tax law. It is a multi-mode survey, with paper as the dominant mode. The survey is offered in both English and Spanish on the web. However, respondents who wish to complete a Spanish paper survey must request one. Over the past six administrations, very few respondents have chosen to complete a Spanish-language survey in either mode. Although Spanish-speakers may choose to complete the survey in English (perhaps with assistance from family or friends), it seems likely that Spanish-speakers are underrepresented. In our most recent administration of the ITB survey, we attempted to increase Spanish-language response by emphasizing that the web survey is available in Spanish. Spanish call-outs throughout the materials informed respondents of the paper survey option, but also pointed Spanish speakers to the web survey. In addition, full web instructions were provided in Spanish as well as English. Our paper will discuss the impact of this approach on the number, mode, and timing of Spanish-language completes, as well as calls received on our Spanish-language customer service phone line.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Increasing Data Quality in Web Surveys: Predicting and Managing Undesirable Respondent Behaviors

Predicting Breakoffs in Web Surveys

Felicitas Mittereder *University of Michigan*

Brady West *Univ of Michigan, ISR*

Due to recent general shifts in survey data collection modes from mail to web, respondents who break off from a web survey prior to completing it have become a more prevalent problem in data collection. Given the (already) lower response rate in web surveys compared to more traditional modes, it is crucial to keep as many diverse respondents in the web survey as possible to prevent breakoff bias, maintaining high data quality and producing accurate survey estimates. As a first step of preventing and reducing breakoffs, this study aims to predict breakoff timing on a question level. We analyze data from an annual online survey on sustainability conducted by the Institute for Social Research at the University of Michigan. This study will make use of survey data, along with rich paradata and accessible administrative information from the sampling frame. In addition to well-known factors associated breakoffs such as answering device (e.g. mobile vs. PC) we investigate previous response behavior like speeding and item nonresponse to predict breakoff probability for each respondent on a question level using logistic regression and survival analyses.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

B.Y.O.P.: Build Your Own Panel

Converting Panelists from Mail Mode to Web Mode in Pew Research Center's American Trends Panel

Nick Bertoni *Pew Research Center*

Online probability-based panels offer a number of attractive benefits compared to the traditional RDD telephone survey. While offering researchers increased efficiency and flexibility in data collection, this emerging shift towards the web does not come without its own set of unique obstacles to overcome. A key challenge facing researchers maintaining these panels is how to cover the segment of the population that does not have internet access. Since its inception a portion of the American Trends Panel members have been interviewed by mail. Having conducted the panel for almost 3 years now, we have observed that this mail component presents two major limitations – it requires a long field period (which makes quick, timely polling impossible) and it severely restricts questionnaire design (extensive skips, fills, randomization are all impractical). To overcome these constraints, we set out to convert the mail component of the panel to web so that we will have an entirely online panel. An important consideration while undertaking such a conversion is the composition of the panelists involved. We know that the characteristics of mail panelists are different than those who participate via web, so converting as many panelists as possible was given great attention to avoid attrition bias caused by panelists failing to convert. This paper discusses the sample composition of those who converted to web and compares that to the composition of those who did not convert. We take a close look at how the conversion affected the representativeness of the post-conversion panel and data quality more generally. These findings should be informative to other panel survey designers looking for ways to cover the non-internet population.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Measurement Issues Surrounding the Provision of Health Care and the Affordable Care Act (ACA)

Understanding and Shaping Health Values and Priorities

Katherine Carman *RAND Corporation*

Anita Chandra *RAND Corporation*

Carolyn Miller *Robert Wood Johnson Foundation*

Matt Trujillo *Robert Wood Johnson Foundation*

The views that individuals hold about health and the factors they think drive health, can influence how whole communities make choices to improve health and wellbeing. If individuals do not see a connection between their health and that of their peers, or if they do not perceive interventions to improve health as important, then their communities and national governments are unlikely to prioritize certain investments. Support for health, and not simply health care policies, ultimately rests on the views that society holds not only about the role of government in promoting health and equity in health but also about how health can be created. In this paper, we investigate how attitudes to government interventions and equality are related to perception of how communities influence health and the types of community policies that should be prioritized. Even if individuals agree that policies to improve health are important, they may disagree about how to improve health. Given these possible links between the role of government, perceptions of how we can create health, and policy choices, it is important to understand the mechanisms that might lead to a relationship between these concepts. In this paper, we use a national survey of 11,555 respondents to investigate how perceptions of the role of government and equity are associated with 1) perceptions of health interdependence and 2) the value placed on policies to improve well-being of communities. Perceptions of health interdependence are measured using a series of 6 questions about community factors that influence health such as community safety and social support. The value placed on policies to improve well-being is measured using a series of 5 questions about support policies such as improved housing and food access. Methodologically, our paper is of interest because we fielded our survey in two probability based internet panels.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

When Push Comes to Shove: Moving Respondents to the Web

The use of mail push to web, email or text invite to web, and CATI to conduct interviews with individuals that exited the Temporary Assistance to Needy Families (TANF) program.

Shelley Osborn ICF International

Rebecca Eaton ICF International

Conducting surveys of individuals that leave a state's Temporary Assistance for Needy Families (TANF) program presents several unique challenges due to the population's unstable housing situation, low literacy, lack of access to high speed Internet, and potential mistrust of the state funding agency and research team. In addition, there is little information about engaging households in which a family member, often a grandparent, receives aid for a child but is exempt from the program's work requirements. This paper presents the results of a pilot test using a stratified random sample of TANF leavers to determine the impact of data collection mode on response rate and data quality. Specifically, 150 records were assigned to one of three data collection strategies: 1) mail push to web with mail survey follow-up, 2) email and/or text link to a web survey, and 3) computer assisted telephone interviewing (CATI). The sample was stratified by three case levels: zero-parent, 1-parent, and 2-parent households. All individuals completing the pilot were provided with a \$20 Visa Gift Card incentive. To evaluate the effectiveness of these methods we will compare response rates, item non-response, and overall data quality for all respondents and for respondents in each of the three case types. The results of the pilot will inform ongoing data collection through June 2017.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Increasing Data Quality in Web Surveys: Predicting and Managing Undesirable Respondent Behaviors

Predicting Survey Breakoff in Web Surveys

Anke Metzler *Darmstadt University of Technology*

Marek Fuchs *Darmstadt University of Technology*

Self-administered surveys lack interviewer support in motivating respondents to complete surveys and to provide high quality responses. Hence, survey breakoff is elevated in Web surveys as compared to other survey modes. Interactive features have successfully been incorporated in Web surveys to compensate for potential increases of item missing, non-differentiation or speeding. However, this has not been achieved for survey breakoff. To provide motivational statements to respondents who are about to abandon Web surveys, it is essential to identify these respondents prior to the actual breakoff. Results of a previous study suggest that decreasing response times are a reasonable indicator for subsequent survey breakoff. However, results also indicate that the false negatives (people who speed up but then stay in the survey) are considerably high. Thus, in this paper we assess a combination of multiple indicators for breakoff such as relative speed, item nonresponse rates, the length of answers to narrative open-ended questions and the degree of differentiation in grid questions to predict subsequent survey breakoff. For analyses reported in this paper, we used four Web surveys among university applicants conducted in 2013 (n=7,395), 2014 (n=5,996), 2015 (n=4,034) and 2016 (n=944). We assessed differences in multiple factors for respondents who completed the questionnaire, for respondents who abandoned the Web survey within the following three pages respectively to the page of interest and for respondents who abandoned the Web survey later in the questionnaire. Descriptive analyses indicate that survey breakoffs differentiate significantly less prior to the breakoff, show significant higher item missing rates and report shorter answers than completes. Further analyses of these factors and a combined measure using multivariate and multilevel analyses allow a more reliable prediction of respondents who are about to abandon Web surveys while at the same time reduces the false negative rate.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

When Push Comes to Shove: Moving Respondents to the Web
Offering a QR Code in Mail Surveys: A Smoother Transition to Online Response?

Chan Zhang *Fudan University*

James M. Lepkowski *University of Michigan*

Lirui He *Jinan University*

Studies have shown that offering additional option of responding online tends to lower the response rates compared to the mail-only design. One hypothesis for this phenomenon is that making a choice between responding via mail or Web can be a daunting task for some people and eventually lead to nonresponse. Another population explanation is that the additional effort required to respond online (i.e., turning on the computer and typing the exact survey link into the browser) can lead to procrastination among respondents interested in responding online, and as a result some of these respondents may never complete the survey. To empirically test those two hypotheses, this study compares two methods of inviting people to respond online: a URL link and a QR code. The experiment is conducted on an area-probability sample of Shanghai residents, the majority of whom are expected to be familiar with both methods. The QR code, in particular, has become widely used in Urban China for mobile transactions. Four experimental conditions vary the response options offered in a mail survey: “paper only”, “paper + URL”, “paper + QR code”, “paper + URL + QR code”. If the complexity of making a choice discourages response, response rates should be lower for condition with more response choices. By contrast, if online response depends primarily on the amount of effort required to access the online questionnaire, the “paper + QR code” condition can be expected to yield a higher response rate than “paper + URL”. If both factors play a role, there will be a more complex pattern of response rates across the four conditions. Findings on response rates as well as respondent demographics across the four experimental conditions will be presented.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Leveraging Contact Strategies and Response Propensities to Increase Survey Participation

Experimenting with Contact Strategies for Areas with Differing Expected Levels of Response in the 2015 National Content Test

Jessica Phelan *U.S. Census Bureau*

This paper examines the results of the strategies tested for optimizing self-response, response rates, and language research as it relates to gaining respondent cooperation in the U.S. Census Bureau's 2015 National Content Test. The nationally representative sample's 1.2 million housing units were stratified into low, medium, and high response areas by combining information about the internet connectivity of a census tract with a model-based predicted response score. Housing units were assigned to one of the nine possible contact strategies being tested where the timing, number, and format of materials included in each mailing differed. All housing units were also assigned to one of three language panels, which provided different types of language support in the invitation letters. This paper describes how certain contact strategies may be more appropriate for housing units in areas that have differing expected likelihood of response. Demographic factors such as the age of household members or the race and ethnicity of household members are considered in the analysis as well. The results will be used in planning for future census tests and for the upcoming 2020 Census.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

When Push Comes to Shove: Moving Respondents to the Web
Driving to the internet: A case study of the multi-mode 'wave' methodology for traditional census collection used for the 2016 Census of Canada
Vince Hartung *Statistics Canada*

Governments around the world continue to look for cost-effective means to conduct their national population counts. While research remains ongoing into other alternatives, until a proven method exists for moving away from the conduct of the door-to-door census, governments will continue to conduct and improve on traditional census collection methods. The Government of Canada has been evolving its census collection processes over the past three census cycles (fifteen years) towards an internet-first self-response strategy with a series of complementary collection prompts and in-person follow-up interactions with respondents designed to achieve target response rates and response quality while minimizing cost. Canada has integrated a program of broad public communications, online questionnaires, a sequence of mail products, telephone voice broadcast, inbound and outbound call centres and field collection activities into a tight sequence of respondent contact activities, also known as 'waves'. The program is enabled and supported by a new national, integrated collection management system. This paper introduces the elements of the formal Census Wave Methodology used during the conduct of the 2016 Census of Canada to orchestrate the sequence of contacts with respondents. It includes a discussion of the drivers influencing the balance of tools and methodologies used in different areas of the country based on geography or respondent characteristics; the challenges of integrating these elements into a cohesive collection process; results of selected elements of the 2016 Census wave methodology; and a look forward to remaining challenges and areas of research.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Measurement Issues Surrounding the Provision of Health Care and the Affordable Care Act (ACA)

How Would Better Knowledge Influence Support for the Affordable Care Act? A Simulation and Experiment

Josh Pasek *University of Michigan*

Tobias Stark *Utrecht University*

Jon Krosnick *Stanford University*

Trevor Thompson *NORC at the University of Chicago*

One popular narrative about the Affordable Care Act (ACA) is that individuals liked most provisions of the law even as they opposed the law itself. Implicit in this claim is the notion that support for the ACA would be higher if Americans simply knew more about it. To test this proposition, we conduct both a simulation study and an experimental intervention to see how a more informed public might differ in their attitudes toward the law. Using a nationally representative sample from GfK KnowledgePanel, we asked respondents whether they thought various provisions were in the law as well as their perceptions of those provisions. We found that more knowledgeable individuals (Democrats, Independents, and Republicans) were more likely to support the ACA if they could confidently identify the provisions that were in the law. Overall, we estimate that an electorate that was confidently knowledgeable about which provisions were (and were not) in the law would have yielded a sizable majority of supporters of the ACA instead of the relatively equivocal actual levels of support for the law. We also provided respondents with information about which provisions were in the ACA and — following this new information — reassessed their reported levels of support for the law. For all groups, the provision of additional information improved perceptions of the law. The results indicate that a more informed public would likely have had a more favorable view of the ACA.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Measurement Issues Surrounding the Provision of Health Care and the Affordable Care Act (ACA)

Impact of the Affordable Care Act on Health Care Affordability and Underinsurance Rates: Results from Three States

Patrick Madden *Market Decisions Research*

Brian Robertson *Market Decisions Research*

Mark Noyes *Market Decisions Research*

The Affordable Care Act (ACA) has provided health insurance coverage to millions and has reduced the rate of uninsurance in the U.S. from 17.6% in 2013 to 9.9% in 2016. The Act included provisions to reduce the total cost of health care, including minimum essential coverage standards, subsidies for lower income families, and out-of-pocket cost limits. Despite these efforts, a growing number are underinsured because of high deductible plans, high out of pocket medical costs, and increasing premiums. Other structural barriers limit access to care. The Commonwealth fund found 23 percent of 19-to-64-year-old adults were underinsured in 2014, nearly double that of 2003. Because of the timing of fielding, the survey was not able to measure the impact of the ACA on underinsurance. To understand the impact of the ACA on underinsurance and access to care, we conducted large, population based surveys in three states before and after implementation of the Act. Post-ACA, uninsurance rates declined to 4.2% in Rhode Island (2016) and 3.7% in Vermont (2014), but remained steady in South Dakota (8.0%; 2015). Individuals underinsured because of deductibles or out-of-pocket expenses increased significantly over this time. Underinsurance rates did not change among adults with employer-sponsored insurance, but health plans purchased through health insurance marketplaces had underinsurance rates significantly higher than those with employer based plans. We also examine the methodology by which underinsurance rates are calculated to determine whether cost and structural barriers to care should also be considered as metrics for the underinsured. This paper provides a case study by examining states that implemented their own state based exchanges and a state that did not. Our conclusion is that while the ACA has provided health care coverage to millions, it has left many without sufficient coverage to access needed care or control their health care costs.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

B.Y.O.P.: Build Your Own Panel

The Effect of Incentives on the Response Rates of Panel Members: Evidence from the Gallup Panel

Audris Campbell *The Gallup Organization*

Jonathan Rodkin *The Gallup Organization*

Jennifer Marlar *The Gallup Organization*

Kirti Kanitkar *The Gallup Organization*

Incentives are often used to increase survey response rates, but do incentives increase response rates in the context of a longitudinal panel? Moreover, do incentives increase response rates for particularly hard-to-reach panel members? As overall survey response rates continue to decline, random-digit-dialing sampling methods are becoming increasingly expensive and panels—especially online panels—present an attractive alternative. Researchers must develop methods to ensure high response rates to panel surveys, especially among populations whose response rates tend to lag behind other groups. The research presented in this paper will contribute to the very limited literature on incentivizing a panel. The study used the Gallup Panel to test whether incentives significantly improve the survey response rates – both overall and over time – for individual panelists. Panel members were randomly assigned to one of three treatment groups: (1) a points system group, in which respondents accrue points redeemable for gift cards; (2) a sweepstakes group, in which one or more respondents is selected to receive a gift card at the conclusion of each study; and (3) a control group, which received no incentive. Of particular interest is whether one or both incentive types increase engagement among those with previously low response rates—including blacks, Hispanics, low-income Americans, and Millennials (those 36 and younger). The study also examines whether incentives increase response rates over time among any of these or other groups, or whether offering an incentive results in an initial bump in response rates which declines as the incentive is offered over time. Answers to this research question are especially relevant for the points system, which allows for respondents to exchange points for gift cards only after completing a number of surveys. The findings will examine response rates by treatment group, as well as by demographic characteristics and past participation history.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Measurement Issues Surrounding the Provision of Health Care and the Affordable Care Act (ACA)

Polling on the Affordable Care Act, Lessons for Future Research

Mollyann Brodie *The Henry J. Kaiser Family Foundation*

Ashley Kirzinger *Kaiser Family Foundation*

Elizabeth Hamel *Kaiser Family Foundation*

When President Obama signed the Patient Protection and Affordable Care Act (ACA) in 2010, it appeared that this major health care reform law had defied the fate of its predecessors. While Americans were split in their attitudes on the legislation – 46 percent of Americans had a favorable opinion of the law and 42 percent had a generally unfavorable opinion of the law – the ACA had garnered enough support from Democratic legislators to pass both the U.S. House and Senate. Yet, more than 6 years later, Americans remain split in their attitudes towards the ACA with the most recent polls finding that about equal shares express a favorable and unfavorable view of the legislation. Meanwhile, the newly Republican-controlled Congress has made repealing the ACA one of their top priorities. Using public opinion polling, this analysis examines the past, current, and future of polling on the controversial ACA and how the lessons learned during the past seven years can help researchers in several ways. First, we argue that while gauging public opinion on the ACA using a single question may reduce measurement error and response instability, it provides a primed partisan-driven response – that will likely not change while the politics of the law remain polarized. Second, we argue that because few Americans are actually experiencing any change from the law and are confused by the complex legislation that seeks to transform various aspects of both health insurance and health delivery, individuals are more likely to rely on their partisan heuristics for their opinion of the law. Regardless of ACA's fate, these lessons will be key for public opinion scholars going forward on how to gauge the American public's opinion on controversial laws in highly polarized political climates.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Leveraging Contact Strategies and Response Propensities to Increase Survey Participation

The Impact of Different Contact Modes on Survey Response Bias

Caitlin E Deal *American Institutes for Research*

Rebecca Medway *American Institutes for Research*

Shazia Miller *American Institutes for Research*

Trey Miller *RAND*

Past research shows that different types of survey reminders are more effective in eliciting response for particular populations or groups of respondents (Dillman et al, 2011). Email reminders may be more effective than telephone prompts for some individuals, while the reverse may be true for others. As a result, surveys often use a variety of approaches to contact sample members when possible. However, these efforts may be hampered by missing contact information (e.g. no email address) or lack of participant agreement to be contacted in certain ways that require prior consent, making it impossible to use these contact methods for all of the sample members. If different follow-up methods are used for different respondents, this use of inconsistent methods can bias the survey results, even if response rates ultimately increase as a result of these contact methods. This study will examine whether consenting to be contacted using particular methods affects survey nonresponse bias using a recent web survey of community college students. Sampled students were asked to consent to receive text or Facebook message reminders as part of a follow-up survey. Seventy-two percent of respondents agreed to be contacted by text, and twenty-five percent agreed to be contacted by Facebook message. Key survey outcomes and demographics will be examined as predictors of the likelihood of consenting to each of these methods. We will determine whether these predictors are different for the two contact methods. In addition, we will assess whether respondents who consent to be contacted by text or Facebook also provide higher quality data overall by looking at item nonresponse and straightlining (choosing the same result to quickly get through the survey). Finally, we will incorporate responses to the follow-up survey to conduct a nonresponse bias analysis.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

B.Y.O.P.: Build Your Own Panel

Developing Random Probability Web-CATI Panels: Evidence from the UK and Beyond.

Kirby Swales *NatCen Social Research*

Curtis Jessop *NatCen Social Research*

Social researchers in the UK continue to be under pressure to produce survey results at a lower cost and in a shorter time period. In this context, the 'Gold Standard' survey approach of the face-to-face interview, random-probability sample can be too slow or costly when compared to faster, cheaper non-random, digital survey approaches. However the non-probability web panel approaches often used in market research, while capable of producing results quickly and cheaply, may not provide the robust estimates needed for social research purposes; recruitment methods may systematically exclude sections of the population (e.g. those without access to the internet) or show bias towards others (e.g. those that self-select to take part, or those that are 'easily available'), and non-random methods preclude the valid use of statistical testing techniques. NatCen, the UK's largest independent social research organisation, has recently completed a feasibility study to setup a random-probability-based research panel, which would provide high quality research. This panel is now operational and will have undertaken over four waves of surveys before the time of the AAPOR conference. In this presentation we argue that such a panel is an important resource for the social research community and outline our approach to developing it. Specifically we reflect on our approaches to panel recruitment, maintenance, and fieldwork, and their roles in creating a high quality panel that bridges the gap between 'gold-standard' surveys and existing cheaper /quicker web panel alternatives. We will also place the NatCen Panel methodology in the context of other international approaches, with the aim of prompting a discussion of the relative and advantages of different approaches.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

When Push Comes to Shove: Moving Respondents to the Web

Got Mail? Drivers of mail-to-online response rates

Valerie Lykes *J.D. Power*

Jay Meyers *J.D. Power*

In an effort to combat dwindling survey response rates, J.D. Power conducted a study using a multi-method approach (i.e., mail to online followed by phone interviews) to determine: 1) the “open envelope” rate, i.e. proportion of potential respondents who opened the invitation package; 2) whether envelope type, pre-notification, incentive, or survey length made a difference in response rate; 3) why respondents took/didn’t take the survey; 4) where drop-off occurred in the “Response Funnel” from invitation to completed survey; and 5) whether the low response rates led to a biased outcome. This study was conducted in two parts in April of 2016. First, almost 24,000 invitations were mailed to qualifying car owners inviting them to take a survey from J.D. Power, a global market research firm with a brand awareness level of 80%. After one week of fielding, phone interviews were conducted with three groups: those who completed the survey (n=2,370), those who were invited but had not yet participated (n=21,562), and a control group of an additional 2,000 vehicle owners. Preliminary results indicate that less than half of those invited by mail to participate in surveys opened the invitation envelope, compounded by attrition that occurred throughout the response funnel. Other initial findings indicate that envelope type can positively influence response rate, but is not cost-effective, and that a pre-notification by phone can improve response rate, likely due to reaching the respondent through a different method (phone) prior to the postal mailing.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

B.Y.O.P.: Build Your Own Panel

Piggy-Backing on Face-to-Face Surveys for Online Panel Recruitment: A 3-Country Pilot

Indrek Soidla *Institute of Social Studies - University of Tartu*

Ana Villar *City University London*

Elena Sommer *City University London*

Didrik Finnøy *Norwegian Centre for Research Data*

Bjørn-Ole Johannesen *Norwegian Centre for Research Data*

Nejc Berzelak *University of Ljubljana*

Slavko Kurdija *University of Ljubljana*

Tina Vovk *University of Ljubljana*

Mare Ainsaar *University of Tartu*

Alun Humphrey *NatCen Social Research*

Online survey panels enable cost-effective and timely data collection. However, obtaining a representative sample can be challenging and expensive: sample frames for probability sampling of the general population usually do not include online contact information, and recruitment thus relies on offline contacts that either suffer from low recruitment rates (telephone and mail) or involve large costs (face-to-face). Random selection of individual respondents within households also poses challenges. Carrying out panel recruitment as part of an existing face-to-face data collection effort could serve as a cost-effective way to obtain high quality data from a probability sample. However, this idea has been met in the past with concerns about potential drawbacks related to additional burden for respondent and interviewer. A pilot study was set up to test this “piggy-backing” approach using the European Social Survey data collection efforts. Participants of the CROSS-National Online Survey (CRONOS) panel were recruited at the end of the 2016 ESS interviews in UK, Slovenia and Estonia. Participation involved completing six 20-minute online surveys over one year. Respondents were offered gift cards for £5/€5 with every new survey invitation. Those without internet access for private purposes were offered an internet-enabled tablet, training on how to use it, and telephone support. This paper evaluates efficiency of this approach by examining recruitment rates across countries and across different subpopulations. We will analyse representativeness of the recruited sample, comparing those who participated to those who did not in important variables from the ESS interview. The initial recruitment rate varies by country—between 51% in Estonia and 61% in the UK; by age—between 80% for respondents younger than 30 and 46% among those older than 60; by gender—54% for women vs 60% for men; and by internet access—50% for those without internet access for personal use compared to 78%.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

When Push Comes to Shove: Moving Respondents to the Web
Leveraging ABS to Conduct a Mixed-mode, Multi-phase Survey

Jennifer Marlar *The Gallup Organization*

Kirti Kanitkar *Gallup*

Manas Chattopadhyay *The Gallup Organization*

Rob Andrews *NOAA*

ABS surveys have many benefits including near perfect coverage of US households, the ability to screen and re-contact, and capacity to offer multiple modes of completion. Many surveys have switched, or are considering a switch, to ABS data collection, making continued research in this area important. Gallup conducted a mail survey experiment, on behalf of NOAA fisheries, leveraging the flexibility of ABS. Households in New Jersey and Florida were sent a mail survey in fall 2016, asking about weather and outdoor activities. All households reporting fishing activity and a sample of households without fishing activity were sent a follow-up survey which asked more detailed questions about fishing activity. The mail design of the follow-up survey limited the complexity of the design of the questions and the amount of information that could be collected. To overcome this limitation, a longer and more complex web version of the survey was created (the preferred version), and households were initially pushed to respond to the follow-up survey via web. The first mailing included a cover letter, instructions for accessing the web survey, and a cash incentive. One week later, household received a reminder postcard with web link. One week after that, all non-responding households received a reminder packet with a cover letter, web access instructions, and the paper survey. The findings to be presented will include a discussion of the overall design and implementation of the survey and the response rates for each phase. It will also explore the number of responses by mode and the demographics of the web versus mail responders. Demographic information from the initial household survey will be used to evaluate the follow-up survey responders and non-responders. The presentation will conclude with a summary of learnings and recommendations for other surveys considering similar approaches.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Obtaining Permission and Consent for Hard to Obtain Health Data

A Method for Achieving High Response Rates in National Surveys of U.S. Primary Care Physicians

Michaela Brtnikova *University of Colorado*

Lori A Crane *University of Colorado*

Brenda L Beaty *University of Colorado*

Mandy A Allison *University of Colorado*

Laura P Hurley *Denver Health*

Allison Kempe *University of Colorado*

Physician surveys are important for health services and policy research. However, many survey studies are limited by low response rates (McLeod et. al. 2013). We used previously described strategies for increasing response rates to conduct surveys within pre-recruited physician networks over a 6-year period. Here, we describe: 1) changes in response rates within the lifetimes of physician networks, and 2) administration methods associated with higher response rates. In an initiative to assess vaccine policy issues among primary care physicians, thirteen surveys were administered between 2008-2013 to pre-recruited, nationally representative physician networks of pediatricians, family physicians and general internists. Each network was active for 3 years and responded to 3-6 surveys without incentives. Physicians received questionnaires by e-mail or mail based on their preference. For each survey, the e-mail group received up to 9 e-mailed requests to complete the questionnaire, followed by up to 2 mailed requests and the mail group received up to 3 mailed questionnaires and a reminder postcard. For six surveys, an additional handwritten envelope with a questionnaire was mailed to non-responders, due to a lower response rate at the end of the usual protocol. Overall, response rates varied between 61-88%. There was an average 18 percentage point decline in response rates between the first survey administered to a newly recruited physician network and the last survey administered two years later. Overall, the e-mail group consistently achieved higher response rates than the mail group (74% vs. 62%). An additional mailed questionnaire in a handwritten envelope succeeded in boosting the final response rate by an average of 11 percentage points. We have shown that our method, which combines several recommended strategies, can achieve favorable response rates over 3 years of network participation.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Obtaining Permission and Consent for Hard to Obtain Health Data

Now that you have completed this survey, can we look at your health records? An examination of agreement to link health records after survey completion

Jeanette Y Ziegenfuss *HealthPartners Institute*

Jennifer MRenner *HealthPartners Institute*

Kayla Dean *HealthPartners Institute*

Casey A Easterday *HealthPartners Institute*

To obtain a complete picture of a population's health and well-being it is necessary to leverage both self-report and administrative health records. The value of either of these data sources is maximized when they can be linked on an individual level, however, health data regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) require patients to authorize researchers to gain access to their records for medical or scientific research. In order to draw meaningful and generalizable conclusions from these linkages, the resultant data must be free of bias—a concern so large that the Institute of Medicine published a caution that required consent would likely lead to biased and invalid results (IOM, 2000). Indeed, research has highlighted systematic demographic and health differences between individuals who are or are not likely to agree to record linkage. In this study, we leverage rich frame data in the evaluation of five surveys conducted by a Survey Research Center within a large Midwestern insurer and medical group to examine how methodological decisions such as mode, incentive and survey topic can impact agreement rates and resultant bias. We also consider the impact of the specific ask to use health data such as the duration and scope of the records requested. The five surveys vary across mode (paper, web and telephone), population, incentive structure and survey topic and offer conclusions about which settings promote bias by comparing demographic and self-reported characteristics between those that agree to linkage and those that do not.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Leveraging Contact Strategies and Response Propensities to Increase Survey Participation

Response Rate Projections for Household Screeners vs. Questionnaires: Can the Same Model Be Used for Both?

Katie Archambeau *NORC at the University of Chicago*

Colm O'Muircheartaigh *Univ of Chicago NORC*

Ned English *NORC at the University of Chicago*

Anna Wiencrot *NORC at the University of Chicago*

In the past, survey completion rates were assumed to be relatively steady across the data collection period. That assumption was soon modified to a simplistic curve, which assumed that completion rates would peak several weeks into production, then taper off until the end of data collection. Recently, researchers at NORC have employed field disposition histories for multiple projects in order to more accurately predict final response rates that can consider mid-project fluctuations in productivity. The National Social Life, Health, and Aging Project (NSHAP) does this by monitoring case dispositions, including whether there has been a refusal, and comparing these cases to the final response status of cases with similar weekly dispositions in previous studies. This application is used to model a final response rate for the current study, permitting more informed case releases and early warning of potential production shortfalls. In this paper, we examine whether these questionnaire response rate predictions can be applied to household screening. This research reports on the use of response rate projections for NSHAP Wave 3, which includes both household screening and a questionnaire. We will look to see how well the model performs in predicting household screener response rates at early points in data collection. Does the model need to be tweaked when moving between household screening and questionnaire completion projections? If adjustments need to be made, how is the screening model similar to the questionnaire model, and where are the differences most pronounced? The NSHAP data will serve our guide to understanding similarities and differences between screener and questionnaire response, and we will use this information to inform the audience on how to best utilize this information on future projects. Our research has value for other field data collection projects that would benefit from early indicators of ultimate production and response rates.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Reducing Barriers Due to Language and Coverage

Adapting a Survey Instrument and Methods to Collect Data from American Indian and Alaska Native (AIAN) Head Start Parents

Maya A Reid *Mathematica Policy Research*

Kathleen Feeney *Mathematica Policy Research*

Michael Cavanaugh *Mathematica Policy Research*

Head Start is a national comprehensive early childhood program serving almost 950,000 low-income preschool-aged children and their families across all 50 states, D.C., U.S. territories, and 150 federally-recognized tribal Head Start grantee communities. However, Head Start evaluations have historically excluded programs serving tribal populations from their samples (Marks and Graham, 2004) given tribal concerns about research and unique protocols in working with sovereign nations. Mathematica recently conducted the first nationally representative study of these tribal programs, reflecting a collaborative effort following tribal best practices. Working with this population involved adaptations to the survey instrument and to methodological procedures, including modifying our recruitment techniques and CATI interviewer training to address cross-cultural understanding and tribal parents' specific needs. These respondents have varying levels of internet access, experience with technology, and exposure to the dominant US culture; our adaptations addressed these characteristics and focused specifically on cultural considerations for AIAN populations. This paper will discuss the approach and considerations in conducting a parent survey with this important population, including:

- Changes and additions to the parent survey instrument to measure key topics and constructs specific to American Indian and Alaska Native families.
- Specialized interviewer training, including a session on cultural considerations of this specific population.
- Lessons learned, measurements of adaptation success, and strategies for identifying areas for adaptation and successfully recruiting and integrating subgroups.

The goal of this paper is to share with researchers the challenges of expanding a study to include tribal populations, considerations for improving data collection, and lessons learned that researchers can apply to their own studies.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Obtaining Permission and Consent for Hard to Obtain Health Data

Collecting Data on Minors in General Population Household Surveys: Who to Interview First?

Darryl Adam Cooney *RTI International*

Matt Jans *UCLA Center for Health Policy*

Interviewing minors in general household surveys poses challenges due to the requirement of parental permission and increased potential for refusals. The challenge is greatest when interviews with multiple household members must be conducted. One approach to reduce overall household burden is modifying the order in which interviews are collected (e.g., adult first v. minor first). We conducted such a manipulation in the California Health Interview Survey (CHIS) to examine the relationship between interview/permission order and data collection yield for minors by allowing interviewers flexibility in the order of collecting interviews. CHIS is a dual-frame RDD survey that samples one adult in each household and, when either or both are present, one teen aged 12 to 17 and child aged 0 to 11 in the same household. After a short household screening questionnaire with any adult over 18 years old, the sampled adult completes an interview for themselves. That same person or a “sufficiently-knowledgeable” parent answers questions about a child. Either adult, as long as they are a parent of the sampled teen can provide permission for the teen to be interviewed directly. Twenty interviewers were trained to attempt interviews in the following order: Adult -> Teen permission -> Child. The remaining interviewers (roughly 100) attempted interviews in the usual order Adult -> Child -> Teen. Interview and call-record paradata were used to determine the actual order of interviews obtained. Initial results show three important findings: First, the experimental protocol was easy to follow by interviews. Second, adjusting for relevant controls in this non-experimental design, in households with both children and teens, teen permission rate was higher in the experimental protocol (66.7% versus 47.1%) and teen interview completion rate was the same in experimental order. Finally, the child interview completion rate was lower in the experimental order (85.4% versus 69.2%).

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Increasing Data Quality in Web Surveys: Predicting and Managing Undesirable Respondent Behaviors

Sequential Prediction of Respondent Behaviors Leading to Error in Web-based Surveys

Adam Eck *Oberlin College*

Leen-Kiat Soh *University of Nebraska-Lincoln*

Increasing uses of technology and advances in the capabilities of computer-based survey instruments enable the collection of a treasure trove of valuable information useful not only for improving our understanding of public opinion but also providing context for interpreting the responses provided by respondents. Much of this contextual information is sequential in nature, capturing the survey response process as it occurs over time. For example, in web-based panels, data is collected across long periods of time describing the participation of respondents over the course of many surveys, possibly revealing patterns of response and non-response (as well as potential indicators of future attrition). Along shorter time scales, paradata tracked during survey responses capture information about how respondents answer questions, including navigation histories through the survey instrument, how respondents enter or change answers to questions, as well as how much time they spend on each action taken in response to the survey. Analyzing these paradata could be influential in better understanding what leads to high quality data collection (e.g., high completion rates and accurate responses) and what leads to errors by respondents (e.g., straightlining, speeding, item non-response, breakoff). In this research, we evaluate how specialized methods for analyzing sequential information from machine learning--namely, recurrent neural networks (RNNs) related to popular "deep learning" methods in computer science--might be useful for analyzing sequential data from computer-based survey data collection. Previous research has focused on how RNNs can predict a single outcome across an entire sequence of paradata (e.g., whether a respondent completes a survey or breaks off early) (Eck et al., 2015). We extend this research to evaluate how RNNs perform when attempting to anticipate problematic behaviors by respondents throughout a response to a web panel survey, such as actions related to erroneous data collection (e.g., straightening, speeding, or item non-response).

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Increasing Data Quality in Web Surveys: Predicting and Managing Undesirable Respondent Behaviors

The Mythology of Data Quality?: Data Cleaning and Bias Reduction

Randall K. Thomas *GfK*

Frances M. Barlas *GfK*

Nicole R. Buttermore *GfK*

Many researchers have argued that, to improve accuracy, we should clean our data by excluding from analyses participants who exhibit sub-optimal behaviors, such as speeding or non-differentiation. Some researchers have gone so far as incorporating 'trap' questions in their surveys in an attempt to catch such participants. Increasingly, researchers are suggesting aggressive cleaning criteria to identify large portions of respondents for removal and replacement. This not only raises questions about the validity of the survey results, but also has cost implications as replacement sample is often required. For this project, we used data from three different surveys that contained items which allowed us to estimate bias, including items for which external benchmarks existed from reputable sample surveys along with actual election outcomes. Survey 1 had 1,847 participants from GfK's probability-based KnowledgePanel® and 3,342 participants from non-probability online samples (NPS) in a study of the 2016 Florida presidential primary. Survey 2, had over 1,671 participants from KnowledgePanel and 3,311 from non-probability online samples fielded for the general elections in 2014 in Georgia and Illinois. Survey 3 was a 2016 national election study with over 2,000 respondents from the KnowledgePanel. We examined how varying the proportion of respondents removed based on increasingly aggressive data cleaning criteria (e.g., speeding) affected bias and external validity of survey estimates. Across studies, while we found NPS had higher bias than the probability-based KnowledgePanel sample, we found that more rigorous case deletion generally did not reduce bias for either sample source, and in some cases higher levels of cleaning increased bias slightly. Some cleaning might not affect data estimates and correlational measures, however, excessive cleaning protocols may actually increase bias, achieving the opposite of the intended effect while increasing the survey costs at the same time.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Reducing Barriers Due to Language and Coverage

Perceptions of inequality: A survey-based pictorial measure suitable for online and face-to-face use worldwide

M.D.R. Evans *University of Nevada - Reno*

M.D.R. Evans *University of Nevada*

Jonathan Kelley *University of Nevada*

Income inequality plays a central role in modern politics, both in developing nations and in rich ones, both in Europe and the USA. Objective data on the distribution of income within and between nations abounds. But little is known about how – or even whether – the public PERCEIVES income inequality in their nation. Even less is known about how – or even whether – those perceptions affect political attitudes, class identification, or subjective well-being. We offer a survey-based pictorial measure suitable for online and face-to-face use throughout the world, a measure we first proposed in 1992. Data have now been collected using it in large national samples in 43 nations between 1987 and 2016. We are thus in a position to document and evaluate it thoroughly. We provide reference norms for the US and some dozens of other nations; show how these perceptions of present inequality are related to perceptions of future inequality and future performance of the economy; and show how all those affect subjective wellbeing and contemporary politics, including voting in the recent US presidential election. Test-retest reliability is respectable ($r=.39$ over a 2.5 year period in Australia, the only country for which we have data), fractionally higher than test-retest reliabilities for perceptions of class conflict. We suggest that a measure of citizens' perception of income inequality would be a valuable addition to many public opinion surveys and further suggest that these pictorial questions would serve very well as that measure.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Reducing Barriers Due to Language and Coverage

Demographic and Contextual Predictors of Racial Identification among Hispanics

Allison R Sullivan *Civis Analytics*

David Shor *Civis Analytics*

Racial identity among Hispanics has long puzzled those who study race and ethnicity in the United States. In addition, many Americans of Hispanic descent have a stronger connection to their country of origin, rather than a pan-Hispanic identity. Thus, it is unsurprising that Hispanics do not report their race/ethnicity in a consistent manner. In this paper, we explore predictors of Hispanic individuals reporting that they are White. We start with a national voter file and flag individuals with a Census-defined Hispanic last name. Next, we survey 13,419 of those individuals regarding their race, where the respondent can only choose one option. We find that age is strongly and inversely associated with the probability a respondent reports being White. In addition, neighborhoods have a relationship with the probability that individual reports being White. These findings underscore the importance of wording in questions regarding race and ethnicity.

May 19th, 2017
10:00 AM - 11:30 AM
Concurrent Session C

Reducing Barriers Due to Language and Coverage

Design with Translation in Mind: How the Census Bureau Developed Multilingual Interfaces for Automated Data Collection Instruments

Brianda Perez *U.S. Census Bureau*

LEP populations face a distinct set of challenges when completing Census questionnaires. The Census Bureau aims to improve service to LEP respondents by translating the Census questionnaires and supporting materials. The 2016 Census Test gave the Census Bureau an opportunity to test the use of automated instruments translated into Spanish, Chinese, and Korean. This presentation will focus on the processes implemented for the translation, development, and testing of translated instruments, challenges encountered, and preliminary language test results.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Behind the Badge: A Nationally Representative Survey of Police Officers (Panel)

Behind the Badge: A Nationally Representative Survey of Police Officers

Rich Morin *Pew Research Center*

Kim Parker

Kim Parker *Pew Research Center*

Renee Stepler *Pew Research Center*

Wesley G. Skogan *Northwestern University*

Andrew Mercer *Pew Research Center*

Chief James Bueermann *President, The Police Foundation*

Fatal encounters between black Americans and police in recent years have fueled a vigorous national debate over police tactics, methods and training. While national surveys reported public reaction to these incidents and the protests that followed, the views of police officers on these issues were unknown except through anecdote or the speculation of advocates on all sides of the debate. To fill this knowledge gap, the Pew Research Center, working with the National Police Research Platform, conducted a nationally representative survey of 7,917 police and sheriff's officers employed in departments employing 100 or more sworn officers. The survey, conducted in the Spring and Summer of 2016, measured officers' attitudes toward the high-profile encounters between blacks and police and the impact these incidents have had on police work. Overall, the survey found that more than eight-in-ten officers say their jobs are harder and riskier now than before these incidents and the protests that followed. Large majorities of officers also say relations between police and blacks in their community have grown more tense and that officers in their departments are now reluctant to carry out some police duties. Officers also view these fatal encounters as isolated incidents and see the protests as motivated largely by anti-police bias. The project also found striking differences in attitudes by race of officer and large differences in the views of police and those of the general public on a range of issues, as measured in a companion Pew Research Center national survey of the general public.

Behind the Badge: Amid protests and calls for reform, how police view their jobs, key issues and recent fatal encounters between blacks and police

Kim Parker

Police say their jobs are harder, riskier and less safe now following a series of fatal encounters between blacks and police. This presentation discusses in detail the major findings of the Pew Research Center Survey and how these views differ by the race, age and gender of the officer, the time that the officer has worked in law enforcement and the size of the city or county where the officer works, among other characteristics. Findings on how officers view issues ranging from use-of-force policies and the use of body cameras to their attitudes on race, how they view relations with minority groups in their community and views on marijuana and gun law reform will also be discussed.

Views from where they stand: The police and the public hold divergent opinions on key aspects of policing and some policy issues

Renee Stepler

The public and police stand on opposite sides of an attitudinal gap across a range of issues. One of the sharpest differences between the police and the public emerges over views on the deaths of blacks during police encounters in recent years and the protests that followed many of these incidents. There is a striking racial divide on the views of these events among both the public and the police. The surveys also focus on work experiences finding that police officers view their own job as more frustrating and less fulfilling than most employed Americans do. When it comes to policies around public safety, police and the public differ on some gun control measures and to what degree marijuana should be legalized. But there are some areas where officers and the public agree. Officers and the public both say that police work is harder today and there is broad support for the use of body cameras. These findings come from Pew Research Center surveys of the police and public, which included a number of identically worded questions to allow for direct comparisons of how the public and police see the role of officers in their communities.

The National Police Research Platform: A Tool to Look Behind the Badge

Wesley G. Skogan

The National Police Research Platform is a nationally representative panel of police and sheriff's departments with 100 or more officers. This presentation will describe in detail what the National Police Research Platform is, how police agencies were selected into the panel and address the steps that were taken to produce a representative sample of law enforcement agencies. of the study sample of agencies..The presentation will also describe the way that researchers can use the Platform to expand knowledge of police performance and culture.

The View from Behind the Badge: What Law Enforcement Agencies Need and Want to Know

Chief James Bueermann

Chief Bueermann, president of the Police Foundation in Washington D.C., offers the law enforcement perspective on the findings of the Pew Research Center survey of police and how agencies and policymakers might use these results as part of efforts to reform police practice.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Race, Tribe and Tribal Enrollment Research for American Indians and Alaska Natives: The Challenges of Measuring a Diverse Population (Panel)

Race, Tribe, and Tribal Enrollment Research for American Indians and Alaska Natives: The Challenges of Measuring a Diverse Population

Rodney L Terry *US Census Bureau-CSM*

Laurie Schwede, Rodney L, Terry, and Leticia Fernandez

Laurie Schwede *U. S. Census Bureau*

Aleia Clark Fobia *U.S. Census Bureau*

Jessica Graber *US Census Bureau*

Among the race groups classified by the U.S. Office of Management and Budget (OMB), the American Indian or Alaska Native (AIAN) category is uniquely defined because it includes the provision of maintaining tribal affiliation or community attachment. The 2010 Census race question had a checkbox for the AIAN category, as well as an instruction to print the name of the “enrolled or principal tribe” in a set of write-in boxes. Previous qualitative research and consultations with AIAN stakeholders have shown that many American Indians, Alaska Natives, and Central/South American Indigenous people find the current tribe instruction confusing or not inclusive of their tribal affiliation, largely because of great diversity in how these groups self-identify their racial and tribal groups. Furthermore, because the resulting data are problematic as a source of specific information on the enrolled AIAN population, various AIAN stakeholders have asked for better data to describe this segment of the AIAN population. To help address these challenges, we present a panel of four 2020 Census research studies conducted to improve race and tribal enrollment measurement of the American Indian and Alaska Native populations. The first two studies examined alternative ways to elicit race and tribe information, using focus group and cognitive interview pretests. The last two studies helped develop and pretest a separate and specific tribal enrollment question. Lastly, we present a synthesis of key findings and implications from all four studies. Topics discussed include respondent preferences of how race, tribe, and tribal enrollment information should be elicited, as well as the political and cultural complexity involved in measuring single concepts among a multicultural, hard-to-recruit, and hard-to-count AIAN population. We identify formats from this research selected for national quantitative tests, and discuss implications of this study and its methods for the wider survey methods community.

Writing in Tribe: Focus Group Pretesting of Methods to Collect Race and Tribe Data with Very Diverse American Indian And Alaska Native Populations

Laurie Schwede, Rodney L, Terry, and Leticia Fernandez

The diverse American Indian/Alaska Native (AIAN) population in the U.S. includes American Indians, Alaska Natives, and Central/South American Indigenous people (CSAI). Prior research indicated that these groups varied in their understandings and interpretations of the 2010 census race question and the instruction to “Print your enrolled or principal tribe.” While American Indians generally understood and reported tribal enrollment, some Alaska Natives had difficulty with “enrollment,” printing names of both corporations and villages, unsure whether the question was asking about enrollment or their racial identity. Furthermore, CSAI participants were not familiar with this concept. Could change in instruction wording improve the clarity and accuracy of reporting across these diverse groups? We present findings

from focus groups testing six alternative race question versions with these AIAN groups. Five versions included the stem question, “What is your race or origin?” paired with alternative AIAN instructions to print your: 1) “enrolled or principal tribe,” 2) “enrolled or affiliated tribe,” 3) “affiliated tribe,” 4) “village or tribe,” or 5) “specific origin.” The sixth alternative omitted “race” and “origin” from the question stem and instead stated: “Which categories describe you?” In 2014, four focus groups were conducted with American Indians and two each with Alaska Natives and CSAI participants. We discuss how participants answered the six alternative race/tribe questions and whether the wording of the stem or instructions of any version inhibited or changed responses compared to other versions. American Indians preferred instructions including “enrolled” and “tribe,” while Alaska Natives and CSAI participants preferred forms asking for “specific origin” or the generic, “Print, for example...” We present question/instruction wording recommendations and the hybrid versions used in a later national test. Finally, we discuss the implications of this race/tribe study and its methods for these and other hard-to-count groups for the wider survey methodology community.

Writing in Race: Cognitive Testing of Experimental 2020 Race and Ethnicity Questions

Aleia Clark Fobia, Rodney L. Terry, and Laurie Schwede

In the 2010 Census, race and ethnicity were measured in separate questions, which past research has found problematic. The purpose of the current research was to test experimental features of three combined race and ethnicity questions, designed to improve race and ethnicity reporting. They combine race and ethnicity response categories in one question, and include a new category for Middle Eastern or North African respondents. Another distinctive aspect of these experimental forms is a write-in line for each response category. The 2010 Census race question had a checkbox for the AIAN category, as well as an instruction to print the name of the “enrolled or principal tribe” in a write-in line. The write-in line is important for American Indians and Alaska Natives since these data are used to distribute federal funding. We report findings from 36 cognitive interviews with respondents from several racial and ethnic groups, including American Indians, Alaska Natives, and Central/South American Indigenous people (CSAI) who speak Spanish. The testing focused on differences in the write-in line instructions for AIAN respondents, as well as the detailed checkboxes for the AIAN category. We found that respondents tended to provide more detail about their racial or ethnic background when the detailed checkboxes were not labeled with specific AIAN groups, but instead were labeled with broad AIAN subgroups. A common theme across respondents was the perception of fairness concerning which groups were specifically named on the form. Overall, respondents preferred broader categories in the detailed checkboxes that then had a write-in line to report specific groups. However, results from a national, split-ballot test in 2015 show that when the broader categories were used, respondents wrote in detailed information less often, resulting in a loss of detailed race data for AIAN groups in comparison to data collected with the 2010 Census form.

Research to Develop a Tribal Enrollment Question for American Indians and Alaska Natives

Rodney L. Terry, Laurie Schwede, and Aleia Clark Fobia

Among the race groups classified by the U.S. Office of Management and Budget (OMB), the American Indian or Alaska Native (AIAN) category is uniquely defined because it includes the provision of maintaining tribal affiliation or community attachment. The 2010 U.S. Census race question had a checkbox for the AIAN category, as well as an instruction to print the name of the “enrolled or principal tribe” in a write-in space. Because data collected from the write-in space most likely include both enrolled and non-enrolled tribal identities, the data are problematic as a specific source of information on the enrolled AIAN population. Various AIAN stakeholders have asked for better data to count the

enrolled AIAN population, as they are the segment of the AIAN population that is eligible for AIAN federal funding programs, and has an administrative relationship with their tribes. In response to these issues, we present findings from 2020 U.S. Census focus group research that helped develop a separate and specific tribal enrollment question for the AIAN population. Eleven focus groups were conducted with AIAN respondents who varied by several criteria, including age, gender, enrollment status, and level of engagement with their tribal communities. We found that (a) enrollment criteria and processes vary greatly by tribe, (b) enrollment status can change, (c) Alaska Natives have complex enrollment relationships that may need multiple questions to measure, and (d) participants thought the best question design is simple, clear, and direct. Finally, we present implications of this research for questionnaire design and the collection of survey data on sensitive topics. These findings informed development of three enrollment question formats that underwent cognitive testing, and are part of focus group, cognitive interview, and national test research to help inform whether a tribal enrollment question is feasible for collecting data in the 2020 U.S. Census.

Measuring Tribal Enrollment among American Indians and Alaska Natives: A Cognitive Pretest of Multiple Approaches

Jessica E. Graber, Anna Sandoval, Rodney L. Terry, Laurie Schwede, and Aleia Clark Fobia

Current methods used by the U.S. Census to measure tribal affiliation among American Indians and Alaska Natives (AIAN) do not distinguish between enrolled and non-enrolled status with a tribe or village. The resulting data cannot be used to specifically describe the enrolled AIAN population, which has led to various stakeholder requests for additional data. As part of continuing research to improve race and ethnicity measurement, the U.S. Census Bureau conducted cognitive pretesting of a tribal enrollment question in 2016 with 64 respondents that included American Indians, Alaska Natives, and Central/South American Indigenous people (CSAI) who speak Spanish. Systematic recruitment from seven locations across the U.S. was conducted to achieve diverse respondent representation, including across federally-, state-, and non-recognized tribe types. Respondents also varied by age, gender, education, language, enrollment status, and level of engagement with their tribal community. Respondents completed three self-administered questionnaires with unique versions of questions designed to measure tribal enrollment status. Probes on key terms and phrases, such as “tribe,” “village,” “community,” “council,” “association,” and “corporation,” were used to determine the comprehensiveness and suitability of each term. Overall, American Indians had the least difficulty completing the enrollment questions in the various formats. Additional findings include Indigenous respondents not understanding the meaning of tribal enrollment because it is not relevant to their experience, and Alaska Natives preferring formats allowing them to report corporation relationships in a separate question. Furthermore, respondents were mixed in their opinions on whether the tribal enrollment data should be collected. We discuss a hybrid enrollment question format informed by these results, which will be tested in a national test in 2017 in preparation for the 2020 Census. Finally, we present challenges of and mitigation approaches to identifying and recruiting eligible respondents from a hard-to-reach AIAN population, and their implications for recruiting other respondent types.

The Quest to Develop One-Size-Fits-All Questions on Race and Tribal Enrollment for the Very Diverse American Indian and Alaska Native Population

Hyon B. Shin, Laurie Schwede, Rodney L. Terry, Aleia Clark Fobia, Jessica E. Graber, and Anna Sandoval

Among the race groups classified by the U.S. Office of Management and Budget (OMB), the American Indian or Alaska Native (AIAN) category is uniquely defined because it includes the provision of maintaining tribal affiliation or community attachment. The 2010 Census race question has a checkbox

for the AIAN category, as well as an instruction to print the name of the “enrolled or principal tribe” in a set of write-in boxes. Previous qualitative research and consultations with AIAN stakeholders have shown that many American Indians, Alaska Natives, and indigenous Central and South Americans find the current tribe instruction confusing or not inclusive of their tribal affiliation, largely because of great diversity in how these groups self-identify their race and tribe groups. In addition, because the data are problematic as a specific source of information on the enrolled AIAN population, various AIAN stakeholders have asked for better data to count this segment of the AIAN population. As part of a panel on measuring race and tribal enrollment for the AIAN population, we present a concluding synthesis of key findings and methodological implications from the presentations on four 2020 Census research studies conducted to improve race and tribal enrollment measurement for the AIAN population. After briefly summarizing findings from all four studies, we identify cultural, political, and survey methodology issues with measuring race and tribal enrollment across three diverse AIAN subgroups: American Indians, Alaska Natives, and Indigenous Central and South Americans who speak Spanish. Among the issues discussed are the need to design clear questions that do not appear biased in favor of any group, as well as research that maximizes participant diversity within AIAN subgroups. We also discuss lessons learned regarding research design, methodology, and recruitment issues that may be applicable to research on other sensitive topics and hard-to-reach populations.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Methodological Brief: Advancing Methods in Emerging Technologies

Geo-sampling and Drones for Estimation of Housing and Households

Karol Krotki *RTI*

Safaa Amer *RTI*

Mark Bruhn *RTI*

Jeniffer Iriondo-Perez *RTI*

Listing housing units in a sampled segment is often a laborious, lengthy, and difficult task, especially when maps are of poor quality or not available. Researchers have recently looked into using GIS and drones to not only map out and sample the geographical units but also to implement a low-cost listing operation to prepare for the final stage sampling, selection of households. Our initial attempt to use drones for listing and estimating housing units and number of households was conducted on a very limited basis in Galapagos Island. That study identified room for improvement in the use of drones and Geo-listing for surveys in developing countries. Furthermore, in that study we did not have access to on-the-ground listings so we were unable to evaluate the accuracy of the listings based on the drone technology. In an attempt to replicate, expand, and improve the methodology in a standard setting of a developing country, we conducted a followup study in Guatemala. In this paper, we discuss the use of Geo-sampling supported by drone aerial imagery to enumerate and sample respondents in randomly sampled Grid-cells. We compare listings, for the same areas, based on drone imagery with those implemented using on-the-ground listing. We also describe improved systems for extracting data from the aerial images to provide more accurate information for updating dwelling unit and household counts. For example, we have made progress in identifying multi-story units and units that contain multiple households.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Methodological Brief: Advancing Methods in Emerging Technologies

Statistical Matching as a Supplement to Record Linkage: A valuable method to tackle non-consent bias?

Jonathan Johannes Ephraim Gessendorfer *Institute for Employment Research*

Jonas Beste *Institute for Employment Research*

Joerg Drechsler *Institute for Employment Research*

Joe Sakshaug *University of Manchester*

Record linkage has become an important tool for increasing research opportunities in the social sciences and is likely to become even more important in the “big data” era. Surveys that perform record linkage are often required to obtain informed consent from respondents prior to linkage – which is not always given. This is also the case for the German National Educational Panel Study (NEPS) Starting Cohort 6 and the Panel Study 'Labour Market and Social Security' (PASS), which both ask for consent to linkage to the Integrated Employment Biographies (IEB) of the German Federal Employment Agency. The major concern regarding this is that non-consent introduces an additional source of potential bias in analyses based on survey and administrative data. One strategy to solve the missing data problem induced by record non-linkage is statistical matching. The missing administrative data of a survey unit is estimated by using the data from a statistically similar individual in the administrative data. To evaluate how effective this method is, we use only the respondents for whom the true links are available and induce a synthetic non-consent that leads to bias. By comparing the true (joint) distribution of all successfully linked respondents to the synthetically biased (joint) distribution and the (joint) distribution after using statistical matching for all synthetic non-consenters, one can assess whether or not the bias has been reduced. Preliminary results for NEPS and PASS show that key variables from the employment biography obtained from the record linked and the statistically matched administrative data are correlated. This correlation is, however, not an ideal way to evaluate statistical matching as a method to reduce non-consent bias in the joint distribution of admin and survey data.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Methodological Brief: Advancing Methods in Emerging Technologies

Mobile web survey in the international setting

Mingnan Liu *Facebook*

Laura Wronski *SurveyMonkey*

Nick Inchausti *SurveyMonkey*

There is no doubt that more people are taking web surveys using their mobile devices. In non-English speaking countries, in particular, more respondents complete online surveys using mobile devices than using desktop or laptop computers. This phenomenon has a huge implication on survey research as different devices are associated with different survey measurement and nonresponse errors, which in turn could affect the survey estimates themselves. Also, data quality, as measured by survey satisficing theory, could also be affected by the survey-taking device. An increasing amount of research has been devoted to examining and comparing survey data collected from mobile and non-mobile devices. This study will expand the existing literature by examining mobile web survey data from six countries, up to three samples within each country, and across three waves of data collection. Specifically, the study was conducted in the U.S., U.K. Australia, China, Brazil, and India. Within each country, data were collected through up to three online panel providers, including SurveyMonkey Audience, Cint, and TapResearch. In addition, for each country/sample combination, three waves of data collection were performed, with each wave one month apart. The samples were independent across the three waves. Respondents could choose the device (as opposed to been assigned to) for taking the survey. The same questionnaire (except minor changes to the demographic questions) was used in the study. In the presentation, we will present whether the mobile web usage different by country and sample. Also, we will show the demographic and data quality differences, if any, between mobile and non-mobile respondents by country and sample. This is the first study of this kind that examines the mobile web in such large scale and cross-cultural perspective.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Methodological Brief: Advancing Methods in Emerging Technologies

Big Data, Big Problems: Overcoming Barriers to Consent for Data Linking

Kyle L Endres *Duke University*

D. Sunshine Hillygus *Duke University*

Steven Snell *Duke University*

In the era of “Big Data” researchers seek to enrich survey data by merging in external data. For example, political scientists link to voter files, economists to data about home ownership and assets, and policy researchers to contextual markers of the respondents’ community. Statistical and computing advances increasingly facilitate such matching, but data linkage raises concerns about privacy and informed consent because the process inherently makes use of respondents’ identifying information. Though requesting consent can introduce bias, in which the individuals who agree to have their responses linked differ in important ways from those that do not consent, institutional review boards and data providers often require explicit consent to link. It is thus critical to identify best practices for obtaining consent to pair survey data with external data. Fielding an original survey with an embedded experiment, we explore respondents’ willingness to allow researchers to use their IP address. While IP addresses are a relatively unobtrusive identifier—we find that respondents regard their IP address as sensitive as other identifiers, including name, address, and birthdate. Furthermore, respondents are relatively unwilling to provide consent to use their IP address. We ask whether we can improve this rate of consent when we demonstrate to respondents that we already have the critical information. Randomly assigning respondents to two versions of the linking request—one which simply asks whether we can use their IP address for linking purposes and a second where we make the same request while also displaying to them their IP—we find that respondents consent at higher rates when they are shown their IP address. Generalizing from this finding, we believe that researchers are likely to achieve higher rates of consent for data linking when they demonstrate that they have the necessary information and simply need consent.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Methodological Brief: Advancing Methods in Emerging Technologies

Alternative Tracking: A First Look at Administering Text Message “Mini-Surveys”

Julie Pacer *Abt Associates*

Kelly Daley *Abt Associates*

The Job Search Assistance (JSA) Strategies Evaluation is examining the effectiveness of different approaches to providing job search services in helping cash assistance applicants and recipients find employment. Conducted in four sites and sponsored by the Administration for Children and Families at the U.S Department of Health and Human Services, a key study objective is to examine both employment rates and “time to employment” as a result of the job search services. The analytic goals make quarterly administrative data from Unemployment Insurance records too coarse and six-month telephone follow-up survey may present problems of recall precision. Moreover, the disadvantaged nature of the population makes tracking for a phone-only survey challenging. Monthly follow-up would enable collection of more frequent, more reliable information on individuals’ employment status as well as changes in their contact information. For this purpose, Abt Associates designed short-term electronic “mini-surveys” administered through a phone text (SMS) message and web survey as a means to gather information on participation in job search services, finding employment, and time-to-employment at more frequent intervals, specifically each month between study enrollment and the six-month follow-up survey. Additionally, these mini-surveys requested that respondents provide updated contact information and served as a regular reminder of study participation until the six-month follow-up survey. Participants are offered \$2 for completing each mini-survey. In this brief, we present preliminary methodological findings from the mini-survey effort. We describe lessons learned in fielding a monthly SMS survey as well as contact rates over multiple months and by mode (SMS versus web), site, and other demographic characteristics. We incorporate additional quantitative data from respondent debriefing questions administered in the six-month follow-up survey to describe user experiences with the mini-surveys.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Methodological Brief: Advancing Methods in Emerging Technologies
Mapping the Meaning of Life: Using Open-Ended Surveys and Computational Methods to Extract the Structure of Subjective Well-Being

Patrick van Kessel *Pew Research Center - Washington, DC*

With public discourse growing increasingly polarized, empirical research on core values and the determinants of subjective well-being – the study of "what keeps us going" - has never been more important. This presentation will explore the results of a large-scale open-ended survey and natural language processing analysis designed to identify and quantify the scope of values, attitudes, experiences, and behaviors that contribute to overall subjective life satisfaction and meaning in life across the population. This approach offers advantages over traditional closed-format surveys, which constrain respondent creativity and can force respondents into predetermined and potentially erroneous typologies – as well as interviews and panels, which are often restricted to small samples that make findings difficult to generalize. This research attempts to bridge the gap between these two methods by combining the depth of an open-format interview together with the breadth of a large-N survey. The resulting corpus – 894 pages of text, encompassing 447,338 words written by 1,904 respondents from across the country – is analyzed using topic modeling and other computational methods, revealing a rich array of themes that range from interpersonal relationships, personal hobbies, and financial security, to more abstract concepts like career goals, spirituality, and moral relativism. Findings indicate that underlying these aspects of meaning is a core dimension of experience versus reflection – from self-actualization, to self-transcendence. After exploring the most prominent topics, how they relate to one another, and how they vary across demographic segments, political and religious affiliations, personality types, and overall levels of well-being, this presentation will conclude with a discussion of the potential benefits such methods offer to broader public opinion research – especially research focused on subjective values and beliefs, the study of which can help shed crucial light on what we have in common with each other, and where we differ.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Methodological Brief: Advancing Methods in Emerging Technologies

SMART System: Survey and Measurement using Avatar and Robotic Technology

Yiran Li *Temple University*

Danfeng Xie *Temple University*

Jeffrey F. Durelli *Temple University*

Carole Tucker *Temple University*

Heidi Grunwald *Temple University*

Li Bai *Temple University*

Background: The use of robots has emerged within healthcare and education settings to provide training and opportunities for social interaction. Robotic systems have been developed for children with Autism Spectrum Disorders (ASD) for education and training with early success. The Robot Operating System (ROS) is a flexible framework that aims to simplify the task of creating complex and robust robot behavior across a wide variety of robotic platforms. In this project we develop a scalable adaptive survey platform using robotic technology called Survey and Measurement using Avatar and Robotic Technology (SMART) in real-time for child health surveys. Methods: The NAO Robot (Aldebaran, France) and survey instruments were programmed to adaptively administer health surveys to children. Our proposed SMART platform will capture and integrate multifaceted survey, meta, and para-data collected via video/audio capture, Microsoft bands, and touch pad sensors. ROS was used to integrate algorithms developed by different disciplines, collecting data including galvanic skin response, heart rate, limb movement, ambient light, etc. into a single integrated common publish/subscribe architecture. In parallel, a virtual robot (computer animated avatars) was developed and provides full integration with the data analytics engine. Pilot data was collected during system development on young children and adolescents with ASD. Results: The survey administration is adapted based on the child's request for breaks and a rule-based strategy for adapting the content and timing of administration. It is developed based on para-data of the child's physiologic responses and response patterns. The results are comparable with NAO robot and the two dimensional avatar. The parents, clinicians, and child respondents reported that the robot system survey was enjoyable and provided more ecologically valid interactive administration. Conclusion: This project is unique in its integration of robotic survey adaptive administration based on a respondent's behaviors and physiologic responses. The successful smart-service system will leverage combined expertise across electrical and computer engineering, physical therapy, survey research and data science disciplines.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

The Survey Combo: Methods for Integrating Surveys and Other Big Data Sources
Combining Survey and "Big Data" to Optimize Research and Minimize Error

Tom W. Smith *NORC at the University of Chicago*

The multi-level, multi-source (MLMS) approach is applied to the 2016 General Social Survey (GSS). The GSS is a full probability, in-person interview of adults living in households in the United States and able to do an interview in English or Spanish. The MLMS approach supplements the data collected in the interview with 1) all available data from sample frame (i.e. Census and postal information), 2) process and observational data recorded by interviewers, and 3) auxiliary data linked to the cases at both the micro- and macro-levels. Auxiliary data includes a large amount of variables from commercial databases and to voting and registration records of the addresses sampled and aggregate level information from multiple sources on such matters a socio-demographics, environmental readings, economic conditions, crime, and politics. The survey and auxiliary data are combined and use both methodologically and substantively. Methodological analysis includes examining non-response bias and substantive analysis includes looking at contextual effects such as the cultural of poverty hypothesis that poor people living in poor neighborhood are the most disadvantages and that local crime rates as well as personal victimization both influence attitudes towards crime and crime prevention.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Who, What, Where and When: Characteristics and Behaviors of Online Responders

Email Versus Text, Computer Versus Smart Phone. A Look at How an Adolescent Cohort Accesses a Web Survey

Joanne Delk *UT Health - University of Texas Health Science Center at Houston*

Melissa B. Harrell *UT Health - University of Texas Health Science Center at Houston*

Matthew Thomas *ICF*

Olivia Saucier *ICF*

Joshua Duell *ICF*

Cheryl L Perry *UT Health - University of Texas Health Science Center at Houston*

Conducting web surveys with adolescents has unique challenges as they are less likely than adults to have or use email, the traditional mode to distribute survey links. Additionally, it is unknown whether they are more likely to take web surveys using a computer, tablet, or smart phone. The Texas Adolescent Tobacco and Marketing Surveillance System (TATAMS) began following a cohort of 3,907 6th, 8th, and 10th grade students in the 2014-2015 school year. Participants were recruited from and took the baseline survey at school, while the follow-up surveys are done via the web. Dependent on available points of contact, participants receive links to follow-up surveys via guardians' and participants' email, text message, and/or conventional mail once every 6 months. The survey is optimized for use on smart phone, tablet, and computer and captures the device type and survey link used (i.e. email, text or conventional mail). Data will be presented from the fourth follow-up survey conducted October through November 2016. To better understand the effectiveness of the various methods of contact we will summarize the device and link participants are using to access and complete the survey, by age, gender, race/ethnicity, socio-economic status, and immediate or delayed survey taker status (i.e. subjects took the survey before versus after the first reminder). We will also explore issues of survey break-off and exiting and reentering the survey as they relate to device and link type. In earlier waves, up to 75% of our adolescent cohort used a smart phone to access the survey, while other studies show that number is typically closer to 40% in young adults. These results can help inform others who conduct web surveys with youth and may suggest shifts in how this new generation accesses web surveys.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Using Pretesting Methods to Develop Key Measures

Trading Precision for Reliability: Time Frames and Estimation Strategies in Answering Questions on Alcohol Consumption

Meredith Massey *National Center for Health Statistics*

Asking questions with a shorter time frame has been shown to result in more accurate estimates of respondent behavior. This is because the longer the timeframe, the harder it is for respondents to remember specific events. However, asking respondents about their behavior within a shorter time frame may not provide accurate estimates of their long-term behavior. Therefore, survey researchers must make trade-offs between reliability and precision when designing questions to measure behavior. This paper discusses the relationship of question time frame to the estimation strategies used by respondents. Findings are derived from a study done by the Collaborating Center for Questionnaire Design and Evaluation Research, National Center for Health Statistics on survey items aimed at measuring behavior related to alcohol consumption. The study was a cognitive interview evaluation of alcohol questions intended for use on the National Health and Nutrition Examination Survey and the National Health Interview Survey. Ninety interviews were conducted. Results show that respondents use different strategies to estimate their alcohol consumption for 30 day versus 12 month recall periods. These strategies and their potential effect on survey estimates will be discussed in more detail.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

The Survey Combo: Methods for Integrating Surveys and Other Big Data Sources

Estimation of county-level prevalence of health-related risk factors utilizing data from multiple data sources.

Laura A Dwyer-Lindgren *University of Washington*

Charlton Callender *University of Washington*

Christopher JL Murray *University of Washington*

Ali Mokdad *University of Washington*

The Centers for Disease Control and Prevention (CDC) conducts a number of ongoing surveys designed to measure health behaviors and health outcomes in the United States. Each of these surveys was designed with somewhat different goals in mind, however, and as a consequence each has specific strengths and limitations. We describe methods for combining across surveys to capitalize on the strengths of each survey in order to conduct county-level analyses of health-related risk factors which would not otherwise be possible. The Behavioral Risk Factor Surveillance System (BRFSS) is a continuous telephone survey with an annual sample size of approximately 450,000 in recent years. The large size of the BRFSS allows for a high level of geographic detail, including design-based state-level estimates and model-based county-level estimates. However, all information collected in the BRFSS is self-reported and there are concerns that respondents may intentionally misreport in response to certain items (e.g., systematically under-reporting weight) and may unintentionally misreport in response to others (e.g., not reporting a health condition which has not yet been diagnosed). The National Health and Nutrition Examination Survey (NHANES) combines household interviews with physical examination of approximately 5,000 respondents per year. The physical examination component, which includes anthropometric measures and collection of biomarkers for various conditions, allows for precise determination of respondents' exposure to health-related risk factors. However, the small size of the NHANES precludes subnational analyses. We describe methods to combine NHANES and BRFSS data in order to simultaneously capitalize on the large sample size of the BRFSS and the precision of the NHANES data collection methods. To demonstrate the utility of these methods, we estimate both high cholesterol (distinguishing between diagnosed and undiagnosed) and obesity prevalence at the county level.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Using Pretesting Methods to Develop Key Measures

Cognitive Interviews for Questionnaire Development in the NSHAP Wave 3 Elder Mistreatment Module

Bernard L. Dugoni *NORC at the University of Chicago*

Nola Du Toit *NORC at the University of Chicago*

Melissa Howe *NORC at the University of Chicago*

Kelly Pudelek *NORC at the University of Chicago*

The National Social Life, Health, and Aging Project (NSHAP) is a longitudinal study of community-dwelling older adults, carried out by NORC at the University of Chicago and funded by the National Institute on Aging. To date, NSHAP includes three waves of data, collected in person at five year intervals during 2005-06, 2010-11, and 2015-16. NSHAP's general focus is on the well-being and social worlds of older adults in the United States. In addition to collecting detailed data on social relationships, physical environment, and health, NSHAP measures elder mistreatment and neglect among community dwelling adults in the United States. This methods paper focuses on the evolution of the NSHAP mistreatment and neglect modules, providing a detailed overview of measurement changes between Wave 1 and Wave 3 as well as scientific and practical rationale for replacing the four main Wave 1 elder abuse questions with the enhanced Wave 3 elder abuse and potential neglect modules. Our discussion includes an overview of how (and why) cognitive interviews were employed with a purposeful sample of new respondents to pre-test proposed elder mistreatment and neglect questions for Wave 3. We provide an overview of final decisions made, rationale behind them, and recommendations for how to leverage the NSHAP longitudinal data to study elder mistreatment and potential neglect. Additionally, we discuss the bases for the decision making process for the use of those recommendations and their effect on the revised instrument, as well as discussing uses of the lessons learned from the cognitive interviews for reports and interpretation of study outcomes in the analytic phase.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

The Survey Combo: Methods for Integrating Surveys and Other Big Data Sources
Data Integration Innovations to Enhance Analytic Capacity and Inform Policy

Steven B Cohen *RTI International*

The quality and content of national population-based economic and health care surveys are enhanced through integrated designs that include the conduct of inter-connected surveys to medical providers, businesses, and medical facilities. Their analytical capacity is dramatically enhanced through data integration efforts that link additional medical, behavioral, environmental, socio-economic and financial content from multiple sectors. Advances in data science are also serving to facilitate the effective and efficient utilization of statistical methods in concert with big data applications to develop these enhanced analytical platforms and infrastructure. There are additional auxiliary and ancillary data resources that could substantially strengthen the capacity to conduct more extensive analyses of social science and health related data. However, in order to address confidentiality provisions inherent to these efforts, data providers are required to de-identify and further restrict individual-level data prior to releasing them for approved projects by masking or removing certain demographic, socio-economic and health content. Consequently, the influence of the measures collected in these supplemental datasets cannot currently be assessed by restrictions in the set of variables available for linkage. This presentation will describe innovative data integration efforts that have been employed to link these restricted data resources with nationally representative socio-economic and health related data using statistical matching and model-based techniques. This will include enhancements to the analytic capacity and utility of clinical trial Project Data Sphere® content through the integration of de-identified and demographically limited patient-level data with nationally representative health related data in the Medical Expenditure Panel Survey (MEPS). The inclusion of the measures obtained from these ancillary datasets will significantly enhance the analytic capacity and utility of the nationally representative data, further stimulating hypothesis generation and the initiation of new studies that explore newly identified relationships.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Increasing Data Quality When Collecting Sensitive Data

Lies in the Fast Lane: Comparing Self-Reported Speeding Citations on Surveys with Speeding Convictions from Driving Records

Randolph Atkins *National Hwy. Traffic Safety Admin.*

Christian Richard Battelle

Betsy Payn Battelle

Justin S Graving Battelle

Ta Liu Battelle

Respondent effects, such as inaccurate recall by respondents and social-desirability response bias, can limit the accuracy of self-report data. In many areas, including voting, health care, and criminal convictions, researchers have used administrative data to validate self-reported behavior. This study examines respondent effects on the accuracy of self-reported survey data on driver speeding citations. An address-based mail survey of licensed drivers in Idaho asked for self-reported information on speeding citations. To assess the validity of these reports, we matched the survey data with actual speeding conviction data from the past three years for the 1925 drivers that responded; 25% of convicted speeders failed to report their conviction from the previous year. An ordinal logistic regression analysis was conducted to validate the accuracy of self-reported speeding convictions after adjusting for demographic and driving exposure factors. Significant predictors of accurate self-reporting included: total speeding convictions in past 12 months, number of convictions of all types, time since last speeding conviction, and the image management component of the Driver Social Desirability Scale. Findings on time-decay effects on memory and social desirability bias are presented and the implications of the results for the validity of self-reported survey questions are discussed.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Examining Voter Turnout in the 2016 Election

Behavioral Influences on Self-Predicted, Self-Reported, and Actual Voter Turnout

Julia Pollak Pardee RAND Graduate School

Many political polls use voters' pre-election estimates of their own likelihood to vote to identify likely voters. Similarly, most studies of voter participation use post-election self-reported vote as the measure of turnout. Yet both self-predicted voting likelihood and self-reported voting tend to differ substantially from official counts. Overconfidence, social desirability bias, and selection bias are two leading explanations for the systematic upward bias. This study examines these and other behavioral explanations, by matching the RAND 2016 Presidential Election Panel Survey (PEPS) to validated voting history data from official voter files and to prior surveys conducted in the RAND American Life Panel (ALP). The 2016 PEPS has a probability-based sample of roughly 2,600 and is a detailed survey on voter intentions and opinions. Prior RAND ALP surveys have measured a range of personal, psychological, and behavioral characteristics—such as subjective wellbeing, decision-making skill in other contexts, and cognitive biases. Using this richer suite of explanatory variables—some of which have not yet been studied in connection to political participation—I explore underlying reasons for observed relationships between voting behavior and demographic or socioeconomic variables. After characterizing the differences between the populations of self-predicted, self-reported and actual voters, I discuss the implications for weighting adjustments to election surveys designed to reflect the population of actual voters.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Targeting Incentive Use

Financial Incentives with Fixed Costs

Raeal Moore *ACT, Inc*
Ben Earnhart *ACT, Inc*
Jeff Schiel *ACT, Inc*

Even though response rates for online surveys are declining (Singer & Cong, 2013), such surveys continue to be highly cost-effective for conducting survey research. Identifying methods to increase online survey participation is imperative. Systematic reviews indicate that financial incentives improve respondent engagement (Pedersen, 2016; Liu, 2011); however, little research has investigated the impact of providing incentives to a predetermined number of survey participants as a means to keep costs fixed. Further, research is limited in how incentive amount might function differently depending on how long it has been since an organization has interacted with a potential participant. This presentation reports on results from an experiment to evaluate the effects of time since last contact and varying amounts of incentives on response rates, costs, and data quality. A total of 44,845 students who took the ACT in 2016 were randomly assigned into a 2 (month student took the ACT: April versus June) x 5 (incentive amount) research design. The levels of financial incentive were \$0, \$3, \$5, \$10, and \$15. Only the first 250 participants in each experimental group were eligible to receive the incentive. The total response rate was 6.2%. Students offered \$10 or \$15 had higher response rates than students offered \$3 or \$5, and both of these groups had higher response rates than those offered no financial incentive. There was no differential effect of incentives for students who took the test in April versus June. The cost per respondent increased as the incentive amount increased. Data quality varied by incentive amount. Providing an incentive did not improve the nonresponse bias by gender, race, and ACT score. Results show that it is not necessary to provide everyone with the incentive for responses to increase. We will emphasize creating stipulations for receiving incentives given survey participation to improve data quality.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Targeting Incentive Use

Impact of Incentive Amount on Adolescent Participation in an Accelerometer-Based Physical Activity Assessment

Kathryn Trundle *Westat*
Andrew Caporaso *Westat*
April Oh *National Cancer Institute*
Terisa Davis *Westat*
Laura Dwyer *National Cancer Institute*
Erin Hennessy *Tufts University*

This presentation will report the results of an incentive experiment that was embedded into the Family Life, Activity, Sun, Health, and Eating (FLASHE) study, sponsored by the National Cancer Institute (NCI). FLASHE included a national sample of dyads of adolescents aged 12 to 17, each paired with a parent. A total of 5,027 dyads were invited to participate in FLASHE. Parents and adolescents were each asked to complete two web surveys — one on physical activity and one on diet. A subsample of 1,690 adolescents was randomized to receive an invitation for accelerometer use. These adolescents were asked to wear a wrist-worn accelerometer for seven days to track their physical activity levels. Adolescents were randomized to receive either \$20 or \$40 upon returning the accelerometer and they were informed of the incentive amount at enrollment. Although much research exists on the impact of incentives on response rates for household surveys of all modes, little research exists on the impact of incentives on enrollment and compliance in studies involving accelerometer protocols. These analyses will utilize the results of the FLASHE incentive experiment and rich frame data to help determine the feasibility of, and optimal recruitment strategy for, a national accelerometer study of youth with no researcher-participant face-to-face interaction. 693 adolescents agreed to participate in the accelerometer arm of the study (41 percent enrollment rate). Although the enrollment rate overall was higher in the \$40 group versus the \$20 group, the difference was not significant. However, 72.7 percent of participants wore the accelerometer (at least one day with greater than or equal to 18 hours of wear time), and adolescents receiving the higher incentive were significantly more likely to have worn the accelerometer than those in the lower incentive group. These results indicate that higher incentive amounts increase compliance with study protocols.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Increasing Data Quality When Collecting Sensitive Data

Where Were You Born? Asking Sensitive Questions in a Non-government Survey

Austin Countryman *Nielsen*

Yelena Pens *Nielsen*

Robin Gentry *The Nielsen Company*

The American Community Survey (ACS) is an ongoing survey that gathers data and produces estimates about small areas (census tracts and block groups) at more regular intervals than the decennial Census. The estimates produced by the ACS are considered to be high quality, stable, and have the benefit of being government sponsored, which legitimizes the need to ask somewhat sensitive questions. Among these questions are those regarding place of birth and entry into the United States for non-native residents. For non-governmental survey researchers, the estimates produced from these questions offer a valuable opportunity to study the populations of various ethnic groups within the United States, particularly Hispanics. Currently, Nielsen uses a vendor as well as an in-person survey for ethnicity and language universe estimates. There has been a desire to move to the more comprehensive ethnicity, nativity and language questions from the ACS since they allow for the use of independent, up-to-date, Census Bureau based universe estimates rather than third party data. To maximize the comparative value between private survey data and the ACS universe estimates, the same questions must be asked of the population of interest. For this reason, Nielsen plans to test the efficacy of selected ACS language and nativity questions in a non-government survey setting, specifically within the Nielsen Radio Diary. Due to their sensitive nature, however, it is necessary to evaluate their impact on quality metrics before implementation. The primary research questions of the test are: will respondents answer the ACS questions, will there be a negative impact to return rates and/or item nonresponse for the added questions, and what is the impact to representation for Hispanics and Spanish-dominant Hispanics? The sample for this test will consist of previous Hispanic diarykeepers, so that their responses to the first survey can be compared against the test results.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Examining Voter Turnout in the 2016 Election

Why People Don't Vote: Variation across Nonvoting Populations and Types of Elections

Heather M Creek *The Pew Charitable Trusts*

Cliff Zukin *Rutgers University and The Pew Charitable Trusts*

Sixty-three percent of the eligible population did not vote in the 2014 congressional elections and it is estimated that 43 percent did not vote in the 2016 presidential election (McDonald, 2016). The rate of non-voting is even higher in primary elections, local races, and special elections. Nonvoters are often studied as a monolithic group, but there are reasons to believe that nonvoters who are registered to vote differ in important ways from nonvoters who are not registered. This study uses an original probability survey designed by the Pew Charitable Trusts to capture the attitudes of infrequent and nonvoters. A national sample of 3,763 eligible voters was conducted between March 25th and April 19th 2016, including 819 respondents who are not registered to vote, and 302 who are registered but say they never vote. The survey includes new questions that encourage respondents to consider the many types of elections and races when they identify their voting frequency. These questions result in a more accurate self-report of individual vote history and give us the opportunity to compare groups that vote frequently with intermittent and rare voters, registered nonvoters, and the unregistered. This research reports the variation across the nonvoting groups, including key differences in the reasoning behind their lack of participation in elections.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Developing Cost Effective Data Collection Decisions and Methods

Do incentives still matter?

Kerry Levin *Westat*

Jennifer Anderson McNulty *Westat*

Jocelyn Newsome *Westat*

Brenda Schafer *Internal Revenue Service*

Patrick Langetieg *Internal Revenue Service*

Saurabh Datta *Internal Revenue Service*

In the past 15 years, the survey industry has experienced a steady decline in response rates (Groves 2011). Incentives are one technique used to counter this trend. Decades of research have demonstrated that, all else being equal, incentives increase participation rates and reduce refusal rates (Singer & Bosarte, 2006). Shettle and Mooney (1999) concluded that incentives in government surveys provide a “decided cost advantage” in improving response rates, without negatively impacting non-response bias, data quality, or respondent good will. The IRS Individual Taxpayer Burden (ITB) survey is an annual multi-mode survey sent to 20,000 individuals in the United States. It measures the time and money taxpayers spend complying with tax law. During the administration of the 2010 ITB survey, an experiment found that offering a \$2 prepaid cash incentive significantly increased response rates. Accordingly, the IRS has included an incentive in the ITB survey in most subsequent years. However, although incentives are effective, they add substantially to costs in a large-scale household survey. In the face of tightening government budgets, the IRS decided to repeat the incentive experiment to explore how much difference incentives make today as compared to six years ago. For the latest administration of the ITB survey, half of the respondents received a \$2 prepaid cash incentive, while the other half received no incentive. In this paper, we will examine the impact of the incentive on response rates. We will also compare the impact of incentives today to the impact incentives had in the 2010 ITB survey. Finally, we will examine any difference in response rates by subgroups that have been historically underrepresented in the ITB survey (e.g., younger adults, low income respondents, and parents with young children).

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Targeting Incentive Use

Selection Criteria for Diverse Groups: How to Incent Efficiently

Kate Williams *Nielsen*

Robin Gentry *The Nielsen Company*

Stephanie Stern *University of Michigan*

The uptick in diversity in the United States re-emphasizes the basic methodological principle that you cannot accurately measure a population without adequately representing the diverse sub-groups which constitute it. Unfortunately, certain demographics remain both hard to find within the general population, as well as challenging to persuade to cooperate once found. In order to represent those groups properly, methodology should adapt to support their recruitment. Using data from a controlled experiment on incentives, we identified and analyzed several sources of frame-based sample indicators that could be used to target enhanced treatments to groups of interest. This allowed us to build a straightforward criteria-based model which identifies pockets of the sample where these sub-groups are more likely to respond to a moderately increased incentive. Therefore, we expect an improved yield for these diverse sub-groups. It also helped to identify some challenging sub-groups where the incentive tested was not effective and where we should consider alternative approaches to increase response. In this paper, we will discuss the impact of criteria on various demographic groups' yield and also examine if this approach produces bias. Further, findings imply how to develop a method to most efficiently assign higher-cost treatments to diverse populations of interest.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Developing Cost Effective Data Collection Decisions and Methods

Testing the Effects of Token Noncontingent Incentives in Follow-up Mailings in a Mixed Mode Survey

Gerry Dirks *Simmons Research*

Paul J Lavrakas *Independent Consultant*

Lisa Lusskin *Simmons Research, LLC*

Beth Ponce *Simmons Research*

Josephine Leonard *Simmons Research*

In the past decade there has been a shift in the preferred recruitment mode associated with the increased use of mixed-mode surveying. Many mixed-mode surveys start with mail contact/recruitment of persons/households who have been sampled from an address-based frame. Following the precepts of Dillman, Smyth and Christian (2014) about the importance of using multiple follow-ups to nonresponders in mail surveys, those conducting mixed-mode surveys often use three or more mailings before ending their attempts to recruit a sampled unit by mail. Some who conduct such surveys then switch to an interviewer-administered recruitment mode (e.g., using telephone when a phone number can be matched to an address and using in-person when it cannot). But this switch is expensive even when only telephone recruitment is used, and it may be more cost effective to allocate extra resources to the mail stages to increase the survey's response rates thereby reducing the need for interviewer recruitment. The Simmons Research company conducts large national surveys of randomly sampled households in the U.S. each year. In their current ABS methodology they do three mailings to nonresponding addresses with a modest noncontingent incentive in the first mailing as well as promising a large contingent incentive, no incentive in the second mailing, and a token (\$1) noncontingent incentive in the third mailing. An experiment was conducted in 2016 which shifted the token incentive to the second mailing for a random subset of households. Our presentation will discuss the results of this experiment, including how the shift in the timing of sending the token noncontingent incentive affected overall response rates, response rates at each mailing and the telephone stage, and how it affected sample representation overall and at each stage. Effects on total surveys costs will be discussed. Furthermore, findings will be broken out for high density minority areas.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Who, What, Where and When: Characteristics and Behaviors of Online Responders

Who Is Responding to Online Surveys—And When?

Laura Wronski *SurveyMonkey*

Mingnan Liu *Facebook*

Erin Pinkus *SurveyMonkey*

All surveys have some potential for coverage bias, whether caused by an improper sample frame, a mode of administration that does not suit the population, or other factors. Online surveys have one inherent coverage bias—they cannot reach a population of respondents that does not have internet access. Online surveys that begin and complete fielding within very brief time frames, though, are further opening themselves up to bias due to non-coverage. Online surveys have the ability to reach unprecedented numbers of individuals instantly, which means that what used to take days now might take only hours. In the field of political polling, this is a great asset, as substantial sample sizes are needed to allow for analysis among smaller and smaller segments of the population (from voters, to female voters, to female voters in a certain Congressional district, and on and on). But, this agility has the potential to introduce or exacerbate coverage bias. For this study, we examine respondents to SurveyMonkey's Election Tracking poll, which had more than 10,000 respondents each week from January to November 2016. We examine the characteristics of respondents by day of week and time of day. In particular, we are interested in demographics (age, sex, race/ethnicity, level of education), political identification (ideology and party identification), and paradata (device type, completion times). Even if it's possible to do a large-scale, representative survey quickly, is it a best practice? This research intends to produce a set of guidelines for conducting timely—yet high-quality—online surveys.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Increasing Data Quality When Collecting Sensitive Data

What Makes A Sexual Orientation Question Sensitive?

Jingwei Hu Joint Program in Survey Methodology, University of Maryland

Survey researchers have long been concerned about social desirability bias that is caused by sensitive questions. Over the past years, substantial research effort of estimating and correcting social desirability bias has gone to such topics as income, vote, alcohol and drug use, abortion, and sexual behaviors. However, little is known with regard to the topic of sexual orientation. This paper aims to assess the sensitivity of a widely used sexual orientation question and to explore the factors that make this question sensitive. The paper draws upon data from a household labor force survey, which was conducted in 2016 by Joint Program in Survey Methodology at University of Maryland using Amazon Mechanical Turk (MTurk), a crowd-sourcing platform. The sexual orientation question was embedded in the survey. Respondents were asked whether they felt comfortable or not after answering the sexual orientation question. Respondents were required to provide information both for themselves and for other household members. In this paper, I will first assess the sensitivity of the sexual orientation question using respondents' direct rating of comfort with the sexual orientation question. I then will assess the sensitivity of the sexual orientation question on the basis of an indirect indicator, the item non-response on this question comparing with other sensitive questions in the survey. Lastly, I will examine the factors that make respondents feel sensitive. Four hypotheses are going to be tested: 1) the more people are educated, the less sensitive they feel. 2) Younger people feel less sensitive than older people. 3) Homosexual and bisexual feel more sensitive than heterosexual. 4) People feel more sensitive when reporting for their own and other household members close to them than for those not close to them. Implications of the results will be discussed.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Targeting Incentive Use

Encouraging Survey Response Among Diverse, Hard-to-Reach Populations

Alicia Harrington *Mathematica Policy Research*

Betsy Santos *Mathematica Policy Research*

Hannah Murabito *Mathematica Policy Research*

The Unemployment Insurance program aims to reduce financial hardships, assist with reemployment, and ameliorate the negative effects of unemployment on the economy. To better understand the extent to which the program meets these goals, the U.S. Department of Labor commissioned the Longitudinal Survey of Unemployment Insurance Recipients (LSUI). Due to the changes and uncertainty associated with job loss, completing surveys with UI recipients poses challenges: job loss can lead to housing changes, disruptions in phone or Internet services, and unpredictable schedules. The LSUI involved two 25-minute surveys timed to capture the experiences shortly after recipients began to collect benefits and when benefits were set to expire, over about nine months. LSUI recipients lived primarily in two distinct California areas with diverse populations and economies. We administered surveys by web and computer-assisted telephone interviewing (CATI). To encourage response through the more efficient modes (web and CATI call-ins), an incentive experiment offered a total of \$30 for completes using those modes, compared with the \$20 incentive for those who completed via CATI call-out. Given the diversity across the two regions, there are lessons to be learned about the best approach to contacting this hard-to-reach population. This paper analyzes whether the differential incentive influenced the mode of completion and whether this impact varied based on characteristics such as demographics or language of completion. We also analyze the impact of this strategy on overall data collection costs. Finally, we recommend the use of differential incentives to limit data collection costs and further ensure survey access for similar hard-to-reach populations.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Who, What, Where and When: Characteristics and Behaviors of Online Responders

Survey Device and Response Quality

Rebecca Powell *RTI International*

Shengchao Yu *NYC Department of Health and Mental Hygiene*

Robert Brackbill *NYC Department of Health and Mental Hygiene*

Ashley Richards *RTI International*

Mai Nguyen *RTI International*

David Wu *NYC Department of Health and Mental Hygiene*

With current technology trends, many respondents will complete a survey on a mobile device. Pew Research Center monitors the growth in ownership of smartphones and tablets, and as of October 2015 the rate of U.S. adults who own smartphones and tablets was 68% and 45% respectively (Pew Research Center, 2015). The American Association of Public Opinion Research (AAPOR) task force on emerging technologies found a relationship between this growth in ownership and the increase in use of mobile devices to complete surveys (Link et al., 2014). Previous research found mobile device users tend to take longer and have higher rates of breakoffs compared to respondents who use computers (e.g., Mavletova, 2013; Bruijne and Wijnant, 2013; Peytchev and Hill, 2010; Antoun, 2015). This means researchers need to understand how response device impacts response quality and what can be done to counteract negative effects. This paper aims to expand on these findings through the use of web data from the fourth wave of the World Trade Center Health Registry. Using both survey data and paradata, which captured the number of backups and device used on each page in the survey, we examine the effects of using a mobile device on response quality. Preliminary data indicates that tablet and mobile device users have higher rates of item nonresponse and rounding compared to respondents who used a computer. In this paper, we expand on these findings and provide further insight into differences in response quality outcomes between devices, adding to the literature on mobile devices.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

The Survey Combo: Methods for Integrating Surveys and Other Big Data Sources

Optimal Integration of Surveys

Mansour Fahimi *GfK*

Increasingly, surveys have to rely on more than one sample source to improve coverage or secure the needed sample size in a defensible and cost-effective manner. Oftentimes, two or more independent samples are selected from separate sampling frames with varying representations of the target population of interest. For instance, one sample could be selected from an online panel, while the second could be a cross-sectional sample from a complete frame. Or, there could be two independent surveys with one relying on dual-frame RDD and another on ABS methodologies. Data integration is also relevant to regional surveys that are conducted independently of national surveys, but in which both surveys collect identical data. In such situations, one might be interested in combining data from a regional survey with those obtained from the corresponding subset of the national survey. In order to increase the inferential base, survey data from multiple samples are combined (integrated) in some optimal fashion prior to analysis. This work provides an overview of the common methods used for combining data from multiple independent surveys, points out the complications and inefficiencies associated with such approaches, and furnishes an efficient alternative that is more stable and computationally less cumbersome.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Developing Cost Effective Data Collection Decisions and Methods

Cost Considerations in Mail Surveys

Mina Muller *Westat*

Pamela Giambo *Westat*

W Sherman Edwards *Westat*

Regina Rudd *Westat*

A recent trend has been transitioning traditionally interviewer-administered surveys to self-administered, either by mail or Web or a combination. Optimal sequences of sample unit contacts in mail surveys have been well-documented, notably by Dillman and colleagues (2014), often using response rate as the outcome of interest. Other important outcomes are the relative cost of each contact, the general representativeness of achieved sample composition, and alleviation of possible bias in key survey estimates. Westat has recently developed and tested mail alternatives to traditionally interviewer-administered surveys, a Companion Survey to the National Crime Victimization Survey (NCVS-CS), and the 50-State Survey of Fishing, Hunting, and Wildlife-associated Recreation. Each of these efforts included pretests and multiple waves of data collection, some including experiments with incentive levels and mailing approaches. The 50-State Survey also includes a large mail nonresponse follow-up study. Since each of these surveys attempts to estimate population totals, so-called “avidity bias” is a concern, i.e., a higher response propensity among crime victims or outdoor recreation participants than among other members of the population. This paper will develop a cost-per-completed-survey metric to compare the effectiveness of different approaches to mail data collection both across and within components of these two surveys. We will compare the effectiveness of different incentive levels, of variations in the mailing package, of different sequences of mailings, and of FedEx versus USPS delivery. Later work will examine what more expensive survey procedures may “buy” in reduced bias.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Examining Voter Turnout in the 2016 Election

A Polling Partnership: NY Times Upshot/Siena College Research Institute in Florida, North Carolina and Pennsylvania

Donald Levy *Siena College Research Inst*

Meghann Crawford *Siena College Research Inst*

New York Times Upshot and the Siena College Research Institute (SRI) partnered to poll and analyzed the data in the presidential, senate and governor's race in Florida, Pennsylvania and North Carolina. Using a list of registered voters, we computed a probability of turnout model based primarily on voting history and designed a stratified sample that considered that turnout probability. Additionally, within the survey traditional likely to vote questions were asked of all respondents. We combined the turnout probability score with the stated likelihood to vote score and applied that combined score as a secondary weight after having raked the sample. This experimental method yielded results in Florida and North Carolina well within the margin of error while the Pennsylvania results were not accurate. This paper more fully described this unique methodology, its strengths and weaknesses, compares the results to those of other pollsters and subsequent analysis of how accurately the turnout model predicted those voters that in the end did in fact vote.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Examining Voter Turnout in the 2016 Election

Early Voters and Late Deciders: Vote Choice by the Timing of the Vote and the Vote Decision

Christopher J Fleury *GfK*

Frances MBarlas *GfK*

Linda McPetrie *GfK*

Annie Weber *GfK*

Randall K Thomas *GfK*

The 2016 presidential election provided many voters with the option of voting before Election Day. At the same time, a sizable share of voters appeared to make their final candidate choice in the closing days of the campaign. The confluence of these developments provides an opportunity to understand the attitudinal and demographic differences between those who took advantage of the opportunity to vote early and those who made their decision shortly before voting on Election Day. In this paper, early voters and late deciders were analyzed using data from a national poll of over 2,000 respondents conducted in the final days before Election Day using GfK's probability-based KnowledgePanel®. Among registered, likely voters, we found that overall 49% reported voting early or by absentee ballot, with 50.3% of Democrats and 45.0% of Republicans reporting voting prior to Election Day. Early voters were less likely than Election Day voters to indicate they decided on the candidate before September. Among the early voters, 70.7% of Democrats and 60.7% of Republicans indicated deciding who to vote for before September. In addition, 19.4% of Early Voter Democrats reported deciding on who to vote for in October or more recently while 29.7% of Republicans reported deciding on their vote choice in the same period. Among Election Day voters, 80.3% of Democrats and 73.6% of Republicans indicated deciding who to vote for before September. Of the Democrats, 11.8% and 16.2% for Republicans indicated deciding in October or more recently. Contrary to our expectations, late breaking developments in the race did not seem to affect Election Day voters as much as it did early or absentee voters. We explore these implications by the various political attitude measures included in the study along with a large number of demographic variables. Results are also compared with exit poll results.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Increasing Data Quality When Collecting Sensitive Data

Evaluation of Classification Error in a Survey on Sexual Assault among College Students

Marcus Berzofsky *RTI International*

Christopher Krebs *RTI International*

Christine Lindquist *RTI International*

This paper evaluates the level of classification error that may exist when surveying college students regarding sexual assault on college campuses. Over the past 2 years, dozens of colleges and universities have conducted surveys attempting to measure and understand the prevalence of sexual assault and the climate surrounding it on their campus. One difference across many of these surveys is how sexual assault is defined within the survey. The definition used can greatly impact how a student responds to whether or not they are a victim. If a student misunderstands the meaning of the definition or does not feel the definition fits their exact circumstances then classification error – a student misclassifying their true sexual assault status – may occur. Given the potential for classification error in this type of survey, the Campus Climate Survey and Validation Study (CCSVS), conducted in the spring of 2015 at nine post-secondary institutions by RTI International and sponsored by the Bureau of Justice Statistics and the Office of Violence Against Women, incorporated multiple indicators in the instrument regarding whether or not a person was victim of sexual assault since the beginning of the current academic year. Using these multiple indicators, this paper shows the results of a latent class analysis (LCA). LCA allows the measurement component of the parameters to be isolated thereby estimating the false positive – an indication a person was victim when their true status is not a victim – and a false negative rate – an indication a person is not a victim when their true status is a victim. Our findings show that a false negative classification error is larger problem than a false positive. Therefore, in the CCSVS, the reported victimization rates are most likely slightly understating the true victimization prevalence at each school.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Using Pretesting Methods to Develop Key Measures

The Development, Design, and Evaluation of a Parent Vaccination Acceptance, Hesitancy, and Confidence Question Set for Multiple Federal Surveys

Paul Scanlon Centers for Disease Control and Prevention

Confidence in, and hesitancy towards, vaccination by parents of young children is an important public health issue. Understanding the reasons behind parents' vaccination decisions, and their extent across the population, allows public health agencies to target messaging and interventions. The Centers for Disease Control and Prevention and other federal and state health agencies monitor parents' attitudes towards vaccination through a variety of surveys, including the three ongoing National Immunization Surveys. Many of these surveys target different populations (i.e. parents of newborns or parents of teenagers), and have therefore used correspondingly different measures of vaccine acceptance, hesitance, and confidence. As part of an ongoing attempt to standardize measures across federal health surveys when possible, CDC determined that a single set of parent vaccination opinion questions should be developed. This presentation will describe the multi-method approach used to develop, design, and test this new question set. Focus groups were conducted with parents of children between the ages of 3 months and 15 years in order to gain a broad understanding of how they made vaccination decisions, including the factors that either pushed them towards or away from vaccinating their children. Combining the focus group findings with a systematic examination of the current measures of vaccine acceptance, hesitance, and confidence, a new preliminary question set was developed that was not targeted towards any one single group of parents. Iterative cognitive testing was then used to refine this question set into three separate short sets of questions (with estimated respondent burdens of 1, 3, and 5 minutes, respectively) that can be used across relevant federal health surveys.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Who, What, Where and When: Characteristics and Behaviors of Online Responders

Examining the Influence of Different Devices on Response Rate and Response Quality of a Web Survey

Anil Kumar Chaudhary *University of Florida*

Glenn Israel *Univ. of Florida*

With surge of internet access, web surveys have become popular among researchers. With the advancement of technology, respondents have access to variety of devices (e.g., smartphone, tablet, laptop, desktop) to complete on-line surveys. Different devices used by respondents cause both challenge and opportunities for survey researchers. This also requires special attention to on-line surveys as most have been designed for desktops and laptops (with minimum optimization for mobile devices). As mobile devices are increasingly used, there is a need to assess how device type influences the response rate and data quality. We used data from the 2015 annual customer satisfaction survey of Florida Cooperative Extension Service (FCES). Specifically, we looked at demographic profile of different device users, how item-response rate, response to check-all that apply question, number of words written for two open-ended questions, total time taken to complete surveys and time spent on a page with two numerical open-ended questions is influenced by use of different devices. We used para-data and screen resolution to differentiate respondents based on type of device used. Among our respondents, 20.7% used smartphone, 10% used tablet, 18.8% used laptop, and 50.6% used desktop to complete our web-survey. Smartphone users were younger and females used significantly higher number of mobile devices compared to males. We found no significant difference in item response rate and response to check all that apply question for different devices. Smartphone users wrote significantly fewer words for one open-ended question compared to desktop users. Smartphone, tablet, and laptop users took significantly more time to complete compared to desktop users. These results illustrate some ways in which device type is associated with respondent attributes and data quality.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Developing Cost Effective Data Collection Decisions and Methods

Impact of Pre- And Post-Incentives on Response Rates to a Web and Mail Survey Using an Address-Based Sample Frame

Kisha Bailly *ICF International*

Kisha Bailly *ICF*

William Bryan Higgins *ICF International*

John Boyle *ICF International*

Naomi Freedner *ICF International*

Providing a monetary incentive for survey completion is an effective tool for increasing response rates. With an address-based sample (ABS) frame and the use of mailed invitations, researchers have the option of utilizing pre-incentives and promised incentives. During the pilot phase of a large, national survey being conducted by web and mail via ABS, we tested various combinations of pre- and promised incentives. Specifically, we drew 2,400 addresses and divided them evenly between 6 different incentive conditions. One third of the sample was assigned to each of three pre-incentive conditions: 1) no pre-incentive; 2) \$1 pre-incentive enclosed; and 3) \$2 pre-incentive enclosed. The sample was then further stratified into two post-incentive conditions: 1) no post-incentive, and 2) \$5 post-incentive offered for completion of the questionnaire. We sent two mailings to the drawn sample. The initial mailing was a paper-based invitation to participate in the survey using a web-based survey (e.g., push-to-web), while the second mailing that was sent 8 days later included both the web site address (URL) and the paper survey package. We calculated the response rate by experimental incentive condition, examined the percent of returned mail overall and by address type, and looked at the median time to complete the survey on the web (in seconds) and on paper (in days). Finally, we calculated the cost per completed survey for each condition.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Examining Voter Turnout in the 2016 Election

How perceptions of social circles shape and reflect voting intentions: a longitudinal analysis

Mirta Galesic *Santa Fe Institute*

Wändi Bruine de Bruin *Centre for Decision Research, Leeds University Business School*

People's perceptions of political preferences of their social circles – their family, friends, and other acquaintances – might both influence and reflect people's own political preferences. We ask two questions: Can people's reports about political preferences of their social circles be used as an additional source of data to increase the accuracy of election predictions? And, do perceived preferences of one's social circle shape one's own voting intentions reported at a later point in time? To answer these questions, we use longitudinal data from the USC Dornsife / LA Times Presidential Election Poll, collected in three waves before the elections and one immediately after. Participants were members of the USC Center for Economic and Social Research Understanding America Study (UAS, a probability-based internet panel). We analyze responses of N=1265 individuals who participated in all four waves, and of N=2229 individuals who participated in the last two waves. We find that social circle reports provide useful information that can inform election predictions on both aggregate and individual levels. They also provide valuable insights about the changes in homophily in social circles of people that supported different candidates, and about the social influences shaping individual voting intentions.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Using Pretesting Methods to Develop Key Measures

Challenges in Conducting a Partial Redesign of the National Survey on Drug Use and Health

Struther L Van Horn *Kent State University*

Jonaki Bose *SAMHSA*

Rachel N. Lipari *SAMHSA*

The National Survey on Drug Use and Health (NSDUH) is an annual survey of the civilian, non-institutionalized population of the United States aged 12 years old or older. Data from NSDUH provide information on illicit drug use, alcohol and tobacco use, substance use disorders (SUDs), substance use treatment, mental health issues, mental health service use, and co-occurring SUDs and mental health issues. For 2015, a number of changes were made to the NSDUH questionnaire and data collection procedures. New data collection equipment was introduced in 2015 to replace the older equipment in the field. Respondent materials also were updated. The partial questionnaire redesign most impacted the prescription drug questions which were redesigned to revise the definition of misuse, shift the focus from lifetime misuse to past year misuse, and collect information about any past year prescription drug use rather than just misuse. Other major revisions included the addition of questions about adult sexual identity and sexual attraction and whether respondents had members of their immediate family in the U.S. military. These changes were intended to improve the quality of the data collected and to address current substance use and mental health policy and research needs. The presentation will first discuss a variety of challenges and processes undertaken for the redesign of a large, nationally representative survey, including: early planning and stakeholder engagement and field tests and methods with small sample sizes (<2,000). The presentation will then discuss the process of deciding trend breaks and analytics, trend breaks that occurred in 2015, as well as how the process is documented for users. The presentation will also briefly discuss the redesign impact analysis based on 2015 data. Finally, the presentation will discuss public user access of NSDUH data.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

The Survey Combo: Methods for Integrating Surveys and Other Big Data Sources

Combining Survey and Administrative Data: The OHSU Health Insurance Coverage Model

Peter Graven *Oregon Health & Science University*

Thomas Meath *3181 SW Sam Jackson Park Rd*

This paper describes the results of a study to understand Oregon's health insurance coverage immediately following the implementation of major policies of the Affordable Care Act (ACA). The primary research objective is to estimate the uninsured rate and track changes to other types of insurance during this pivotal period. The methodology uses a novel approach, called the OHSU Health Insurance Coverage Model, to combine administrative data from Medicare, Medicaid, and state insurance division enrollment data with population estimates and survey estimates of the uninsured to arrive at more timely estimate of coverage. The 2015 coverage estimates provide levels consistent with enrollment source data. Additionally, through comparison with survey-based approaches the estimates provide some indication of survey response issues that may be occurring. While the model measured a one-time drop in the uninsured from 14.5 percent in 2013 to 5.1 percent in 2014 and holding at 5.8 in 2015, the American Community Survey (ACS) showed a gradual decline from 14.5 to 9.1 to 7.0. The cause of the difference is suggested by the relatively lower count of Medicaid found in the survey. The model showed an increases of 10 points in Oregon's Medicaid program, the Oregon Health Plan. Comparatively, the ACS showed an increase of only 1.3 points in 2014 and another 1.3 points in 2015. The model also calculates that group coverage decreased between 2013 and 2015 by 4 percentage points from 46 to 42 percent, a 9 percent drop. The percentage drop in coverage is largest for the small group market (18%). Given the potential for change in health insurance policy from a new administration, understanding the value of this kind of model is relevant for methodologists and policymakers.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Who, What, Where and When: Characteristics and Behaviors of Online Responders

A longitudinal online study with prospective homebuyers: Who drops out?

Alycia Chin *Consumer Financial Protection Bureau*

Mick Couper *University of Michigan*

Dustin Beckett *CFPB*

Despite price variation for mortgages, many homebuyers do not comparison shop. To study comparison shopping, we designed a longitudinal online experiment with prospective US homebuyers who had accounts on a real estate website. Consumers were randomly assigned to receive shopping advice using 1) messaging, 2) online resources, or 3) neither (control). We explore attrition from that experiment. Participants were asked to complete a baseline survey and up to six check-in surveys online with invitations sent by email. We had 71,600 participants start the study, of which 19,500 were eligible and 14,000 finished the baseline. The baseline collected demographic and psychological characteristics, expectations regarding home and mortgage outcomes, and other variables. The check-in surveys were delivered every two weeks for three months unless participants missed two surveys or bought a home. These check-in surveys focused on recent home and mortgage search activities. In an attempt to reduce attrition, we randomly assigned participants to a 4 (weekday) x 2 (scheduling: yes/no) design. Participants were assigned to receive survey invitations between Monday and Thursday. Those in the scheduling group could override the default and pick a convenient day. Our analysis models attrition using demographic, psychological, and study characteristics. Overall retention was 60%. The majority (about 80%) of participants who dropped out skipped consecutive surveys rather than returning to the study. We find lower attrition rates for participants with certain 1) demographics: married, white, younger, more educated, and higher self-reported credit score; 2) psychological characteristics: higher levels of patience, subjective numeracy, and financial literacy; and 3) study variables: those in the control group (versus messaging group), those who enrolled in text message reminders about surveys, and those who answered surveys on a computer (versus mobile device). We find no benefit in retention from allowing participants to schedule surveys once controlling for day-of-week effects.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Increasing Data Quality When Collecting Sensitive Data

The Effect of Using a Computerized Voice to Ask Sensitive Questions in the National Survey on Drug Use and Health

Gretchen McHenry *RTI International*

Devon Cribb *RTI International*

Emily Geisen *RTI International*

Joel Kennet *SAMHSA*

Dicy Painter *SAMHSA*

Gilbert Rodriguez *RTI International*

Vorapranee Wickelgren *RTI International*

Hilary Zelko *RTI International*

Beginning in 2015, the National Survey on Drug Use and Health (NSDUH) implemented Text-To-Speech (TTS) software to read the Audio Computer-Assisted Self-Interview (ACASI) portion of the questionnaire, which covers sensitive topics such as alcohol, tobacco and drug use, health issues, and criminal behavior. Before 2015, these items were asked via ACASI but with human voice recordings. A small body of research has shown mixed results on whether use of synthesized speech affects data quality, particularly for sensitive items (Couper et al 2004). A recent experiment found that respondents preferred a human voice, particularly one that sounds more 'empathetic,' but that the type of voice, TTS or human, did not impact respondents' answers to sensitive items regardless of their voice preferences (DiLorenzo 2015). With the recent change in the NSDUH design, there is an opportunity to compare identical sensitive items between 2014 and 2015 to assess potential differences in responses between those responding to a human voice recording and those responding to a TTS voice. Items that did not undergo any other change between 2014 and 2015 will be compared for item nonresponse, frequency of "don't know" responses compared to refusals, and frequency of endorsement of sensitive items. . Sensitive items will include alcohol, tobacco, and illicit drug use and mental health measures. Analyses by demographic categories may help to tease out differential effects, and comparison of sensitive to non-sensitive items might further refine any findings. The results of this research will contribute to the growing body of literature on text-to-speech and other audio software's impact on data quality.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Targeting Incentive Use

Reengaging Respondents after Receipt of an Incentive to Reduce Item Non-Response: When is the Best Time to Reengage?

Jaimie Grazi *Mathematica Policy Research*
Andrew Hurwitz *Mathematica Policy Research*
Martha Kovac *Mathematica Policy Research*
Erin Panzarella *Mathematica Policy Research*

The use of incentives as a strategy for improving item response is well documented in the survey research literature (Brown et al., 2016). Zukerberg et al. (2007) and Brick et al. (2004) suggest that mailing an incentive can increase survey completion success. However, consensus concerning optimal timing for reengaging respondents has yet to be reached (AAPOR Task Force on Survey Refusals, 2014). In this study, we report on findings from a protocol where interviewers reengaged 83 respondents by telephone at various intervals (length varied between 7 and 36 weeks with an average of 18 weeks) in an attempt to improve item response. Respondents in this sample were part of a large multi-site study of young mothers from underrepresented groups who had recently given birth to a child. Participants were asked to complete a brief survey that included sensitive items (e.g., Social Security number). Those who responded were sent a small token of appreciation. Participants who partially completed the survey were classified into two conditions, and both conditions received the same incentive amount by mail as those who fully completed the survey. The goal of recontacting the respondents after sending the incentive was to obtain missing sensitive items that the participants initially refused to answer. Findings from this study suggest that mailing an incentive to survey respondents can increase cooperation and improve item-level response with minimal effort, and that the timing of the incentive mailing can affect item-level response. Central tendency statistics and significance test results will be reported by condition as well as by timing of reengagement. Additionally, a cost benefit analysis will be presented.

May 19th, 2017
1:45 PM - 3:15 PM
Concurrent Session D

Developing Cost Effective Data Collection Decisions and Methods

Mail Survey Experiments: Reminder Postcard and UPS Mail Innovations Envelope

Joseph McMichael *RTI International*

Melissa Helton *RTI International*

Jamie Ridenhour *RTI International*

This paper presents the results of two experiments conducted with a national survey that primarily targeted households in rural counties. With a sample of 63,000 addresses drawn from an Address Based Sampling (ABS) frame we use a two-way factorial design to test whether a reminder postcard significantly improved response rates and whether a UPS Mail Innovations Envelope compared to a 9x12 white envelope improved response rates. We will evaluate the tradeoffs between costs and response rate improvement. We discuss some of the advantages and operational considerations of using UPS Mail Innovations envelopes. We will also compare demographics and key outcomes of interest across the four treatment groups assessing the potential of non-response bias.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Tolerance of Polygamy: A Matter of Portrayal, Parasocial Processing, and Parasocial Comparison

T. Phillip Madison *University of Louisiana - Lafayette*

Tawfiq Abdullah *University of Louisiana - Lafayette*

Television featuring marginalized groups may increase viewers' tolerance toward those groups (Schiappa, Gregg, and Hewes, 2006). This study explored attitudes in response to different portrayals of polygamists, while filling research gaps regarding the role of social comparison in parasocial processing (PSP). We hypothesized those exposed to negative portrayals would develop negative attitudes toward polygamy, while those exposed to positive portrayals would develop positive attitudes. We also asked how those exposed to "mixed" portrayals of polygamy would respond. Finally, we explored social comparison's role in PSP. Researchers administered a pre-test several weeks before a viewing experiment, in which participants viewed 40-50-minute episodes of the HBO series *Big Love*, specially-edited to create positive, negative, and mixed portrayals of polygamy. Researchers immediately administered an identical post-test after viewing. Measures included Attitudes Toward Polygamists (ATP), adapted from Herek's (1988) Attitudes Toward Lesbians and Gays scales, PSP, adapted from Schramm & Wirth (2010), and parasocial comparison (PSC), adapted from the INCOM by (Buunk & Gibbons, 1998), along with demographic variables. Our control group viewed a nature show irrelevant to polygamy. Experimental results follow Schiappa, et al.'s (2005) Parasocial Contact Hypothesis; positive portrayals of polygamists yielded higher ATP scores while negative portrayals scored lower. People exposed to the mixed condition also scored lower on ATP. The control group yielded no significant attitude changes. Finally, higher levels of cognitive, affective, and behavioral PSP were significantly associated with higher levels of PSC. This study demonstrates 1) different portrayals of marginalized groups can have a profound impacts on viewer' ATP, and, 2) PSC is a vital component of PSP, as suggested by Hartmann and Schramm (2008). We provides valuable theoretical additions in understanding the roles of portrayals of polygamy, PSC within PSP, and how PSP affects tolerance of underrepresented groups.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Evaluating Non-probability Samples: An Index of Sample Representativeness

Hee-Choon Shin *NCHS/CDC*

Jibum Kim *Sungkyunkwan University*

The main objective of sampling is to obtain a representative sample for an unbiased and efficient estimate within a budget constraint. A population characteristics or measures of N population elements could be interpreted as a vector on N -dimensional space. Similarly, a sample characteristics or measures of n sample elements could be interpreted as a vector on n -dimensional subspace, imbedded in N -dimensional space. The length of a population vector is defined as the square root of the sum of squares of N components. The length of a sample vector is the square root of the sum of squares of n components. A weighted length of a sample vector could be obtained by weighting the n components with sampling weights. We can measure the sample representativeness as a ratio of the weighted length of a sample vector to the length of the population vector (2016). The current paper is to evaluate the proposed index using simulated data. The 2015 National Health Interview Survey will be used to generate various data sets simulating non-probability samples.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Fielding Section 508 Compliant Web Surveys for Respondents with Disabilities

Bob Davis *Davis Research, LLC*

When web surveys are programmed, much effort is placed in creating a user interface experience that is efficient and engaging for sighted users. This often comes at the cost of making these surveys unusable by those with visual impairments. In addition to being a good practice to design surveys which maximize participation, it may be required by law. According to Section 508 of the Rehabilitation Act of 1986 electronic and information technology developed, procured, maintained, or used by federal agencies is required to be accessible to people with disabilities. This requirement is known as 508 Compliance. An ongoing survey designed for 508 Compliance which Davis Research fields among United States Veterans will be used as a backdrop to discuss best practices and considerations when fielding web based surveys with respondents with disabilities, particularly those who have some form of vision loss.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Public Opinion and the Politics of Education: Convergence in an Era of Polarization

David M Houston *Teachers College, Columbia University*

Perhaps the single most prominent trend in contemporary American politics has been the surge in polarization over the past two decades. On a large array of issues, Democrats and Republicans have grown increasingly distinct. The typical portrayal of federal policymaking is one of excessive gridlock, dysfunction, and the inability to find common ground. Yet it is in this context that two of the most significant federal education laws, the No Child Left Behind Act of 2001 and the Every Student Succeeds Act of 2015, were passed in an overwhelmingly bipartisan fashion. Although public feuds over educational issues may be rhetorically red in tooth and claw, federal education policy thus far has largely escaped the pattern of polarization that has embroiled lawmaking in other arenas like healthcare, gun control, and immigration. Why has education been spared this fate? This study asks whether parents—perhaps the single most prominent stakeholders in educational debates other than teachers—hold policy preferences that enable a less polarized legislative environment. Using nationally representative public opinion data collected by the Program on Education Policy and Governance at Harvard University, I conduct an analysis of the extent to which being a parent of a school-aged child reduces the pull of partisanship and ideological abstraction on a series of contemporary education policy issues, including educational spending, charter schools, vouchers, teacher salaries, and the use of student standardized test scores as a component of teacher evaluation. I find that there appears to be a moderating influence of parenthood on the relationship between political identity and support for a variety of education policy issues. Insofar as there is a consensus on the issues of the day in the educational arena, it is even more pronounced among the parents of school-aged children.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Differential Item Functioning and Non-Health-Related Factors Due to Age in Self-Ratings of Health: Evidence from the Survey of Health, Ageing and Retirement in Europe

Patrick Lazarevič *TU Dortmund University*

Comparative analyses frequently base on respondents' self-rated health (SRH), assuming that SRH is comparable between different groups of respondents. However, this comparability is questionable due to manifold non-health-related factors (NHRF) such as age- or gender-specific expectations for ones' health and frames of reference, or cultural contexts. If SRH is examined, researchers often distinguish between two components of the respondent's answer: latent health and reporting behavior. While latent health exclusively refers to a person's health status, reporting behavior results from a wide variety of NHRF such as general optimism, education, the interviewer, or cultural factors. The analyses of this paper base on the information of 14,400 participants of the fifth wave of the Survey of Health, Ageing and Retirement in Europe (SHARE). First, we predicted SRH via linear regression using a latent health model comprising a wide range of health factors. Second, we used a reporting behavior model covering NHRF to explain the residuals of the first regression. To identify differential response behaviors, all analyses were carried out separately by gender and three age-groups (50-64, 65-79, and 80-90 years). The latent health model shows only minor differences between men and women in explaining SRH, although almost all effects are significant regardless of gender. The amount of explained variance is high regarding all models (with a mean of 46%). The NHRF-model identified consistent influences of interviewer's SRH, respondent's life satisfaction, and country of residence on SRH, while other factors differed by gender and age-group. The amount of explained variance due to reporting behavior in all models can be considered meaningful (with a mean of 9%), given the amount of explained variance by health-related variables. These results illustrate the importance of taking factors such as gender- and age-specific response behavior and NHRF into account while utilizing SRH.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Creating a Comprehensive Database of Medical Marijuana Dispensaries in Los Angeles County

Megan Zander-Cotugno *RAND Corporation*

Eric Pedersen *RAND Corporation*

Elizabeth D'Amico *RAND Corporation*

We have followed a cohort of youth for 10 years who began participating in a voluntary, after-school intervention program during middle school in Los Angeles County. The sample is very diverse: 53% female; 42% Hispanic, 24% Asian, 6% Black or African-American, and 9% other non-white race. We have published papers on a wide variety of topics related to adolescent substance use: differences between racial and ethnic groups; influences of social factors, perceived norms, and neighborhoods; effects on sleep; media exposure and the effect of advertising; and many others. We are now interested in how proximity to and density of medical marijuana outlets near youths' addresses may be associated with marijuana use and consequences. Since there is not a reliable, single source from the State of California or from Los Angeles County that contains accurate, updated information on medical marijuana dispensaries (since most of these stores are currently unregulated), we set out to geocode all medical marijuana dispensaries in Los Angeles County beginning in October 2016. We are using established observational methods based on work from RAND researchers who geocoded dispensary data in Colorado. These methods are designed to give us the most accurate and updated information about stores selling medical marijuana. This poster will review our methods, coding, and verification strategies. If California passes Proposition 64 (legalization of recreational marijuana use for those over age 21) in the November election, more stores will likely emerge and the state will start issuing licenses. At that point, we will reevaluate our methods and update as needed, but methods used in this initial effort will prove invaluable for identifying and locating those stores in the future.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

What Month Did Your Child Receive an Influenza Vaccination? Remembering When.

Tammy A. Santibanez *Centers for Disease Control and Prevention*

Yusheng Zhai *Centers for Disease Control and Prevention*

Lin Liu *NORC*

James A. Singleton *Centers for Disease Control and Prevention*

Background: The National Immunization Survey-Flu (NIS-Flu) provides influenza vaccination coverage estimates for children 6 months-17 years in the United States, based on parental report of their child's vaccination status. Respondents are asked whether the child had received an influenza vaccination and what month it was received. Month of vaccination is used in the Kaplan-Meier estimation of influenza vaccination status, because interviews for NIS-Flu occur during the period of influenza vaccination. For approximately 1/5th of the parentally-reported vaccinated children, the parent does not remember the month of vaccination, which is imputed. **Objective:** Determine 1) characteristics associated with missing month of vaccination and 2) concordance of parent versus provider reported month of vaccination. **Methods:** Data from the 2014-15 NIS-Flu will be examined. The percentage missing month among parentally-reported vaccinated children will be examined overall, by month of interview, by race/ethnicity, age group, survey component, and other variables. A subset of the NIS-Flu cases will be matched to provider records obtained from the NIS and NIS-Teen surveys to compare parental and provider reported month of vaccination. **Results:** Of n=128,143 completed interviews, n=68,982 had a parental report of receiving at least one influenza vaccination. Of these children, 13,057 (18.9%) had parents who could not remember the month of vaccination. Preliminary results show that the percentage with missing month among vaccinated children varied by month of interview, race/ethnicity, and several other factors. By month of interview, the percentage missing month of vaccination were: 17.0% for October interviews, 11.8% November, 12.4% December, 15.1% January, 18.7% February, 20.6% March, 22.5% April, 23.6% May, and 25.9% June. **Conclusions:** Some parents have difficulty remembering their child's month of influenza vaccination, with this difficulty increasing over the course of the flu season. A better understanding of factors associated with missing month responses can help with influenza survey improvements.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Using Paradata-Based Key Performance Indicators to Monitor Implementation of a Split-Ballot Experiment

Peyton Craighill *U.S. Department of State*
Aneta Guenova *US State Department*

Key performance indicators (KPIs) derived from paradata can be used to monitor the quality of data collection. This presentation will describe how KPIs can be used to monitor a split-ballot experiment, highlight continuities between our findings and paradata-related best practices and outline a program for further research. The presentation will focus on the use of a set of KPIs in a split-ballot experiment embedded in a CAPI survey of 1,500 adults. The experiment asked parallel scale questions of each subsample. Respondents were randomized to two different versions of the test questions. Randomization was carried out within each sampling point, with the two questionnaire versions alternated. The maximum number of interviews per sampling point was eight. The researchers identified several KPIs to monitor during data collection: overall number of interviews per sampling point; number of interviews for each questionnaire version per sampling point; date and time of each interview within sampling point to confirm questionnaire versions were alternated. In addition, there were separate variables for stratum, PSU number, sampling point, original or replacement sampled unit, interviewer ID and characteristics, supervisor ID, and length of interviews. Due to budget limitations, it was impossible to ensure daily or weekly monitoring of the data. Instead, the researchers identified two points during fieldwork to conduct reviews of the survey paradata. The first review took place after the completion of a total of 630 interviews across the two subsamples and the second after the completion of 1,270 interviews. After the second review deviations from the survey protocol were identified in about 20% of the sampling points, requiring remedial action. The presentation will detail the data-quality checks that were conducted at each review, describe the process by which remedial action was deemed necessary, and propose a program for the monitoring of fieldwork through paradata analysis.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

The Impact of Delayed Incentives on Future Survey Response

Jordon Peugh *SSRS*

Austin Countryman *The Nielsen Company*

Robin Gentry *The Nielsen Company*

Yvonne Shands *SSRS*

Carrie Skinner *SSRS*

Kate Williams *The Nielsen Company*

There is little research on the impact of the timeliness of a promised incentive or other anticipated follow-up communications on ongoing participation in a study. A large ABS direct mail survey provides an opportunity to examine the impact of delays in mailed promised incentives on participation in future phases of the research. In the first phase of this study, respondents who return completed screeners by mail are in turn mailed a promised incentive (\$5 or \$10). This incentive is mailed on average approximately 11 days after receipt of the screener. In order to deliver incentives quickly (with the goal of keeping response rates high), promised incentives are mailed out before all screener data are processed. As a result, a small portion (2% of the approximately 240,000 incentives mailed out during a six-month study time period) of mailed incentives are returned undeliverable. The vast majority (about 99%) of undeliverable incentives are re-processed and re-mailed with updated information about the respondent. This poster will present data on what happens next with respondents for whom the incentive was returned undeliverable. We will examine the correlation between undeliverable incentives and cooperation with the next phase of the study. We will also control for demographic and other factors that could be associated with both experiencing an undeliverable incentive and participation in future research. Finally, we will present suggestions for balancing the desire for fast delivery of promised incentives with the need to ensure that updated/accurate mailing information is available.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Comparison of Survey Response and Sampling Bias by Sample Frame

Carol Pierannunzi *Centers for Disease Control and Prevention*

Fang Xu *Centers for Disease Control and Prevention*

Pranesh Chowdhury *Centers for Disease Control and Prevention*

William Garvin *Centers for Disease Control*

The Behavioral Risk Factor Surveillance System (BRFSS) is a state-based system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. Since 2011, the BRFSS has conducted interviews by both landline and cell phone, using two sampling frames. Response rates differ by frame and state. This analysis combines BRFSS 2011-2014 data from all states to examine sampling bias based on age, race, sex, education, and healthcare coverage among respondents age 18 and 64 using American Community Survey (ACS) data as a standard. The Mean Squared Error (MSE) of unweighted and weighted samples are compared against ACS estimates by frame and reported telephone usage (landline only, dual user or cell phone only). Results indicate that the landline MSE has increased over time, and cell phone MSE has decreased. MSE also differs greatly by demographic characteristics within sampling frame, usually illustrating that the two samples have unweighted bias in contrary directions: Cell phone samples are biased toward younger respondents and landline samples are biased toward older respondents. As expected, weighting controls for much of the variance of MSE by demographics, which is also corrected by the combination of data from the two sampling frames. As the percentage of cell phone respondents in the BRFSS sample increases, unweighted MSE has declined for most variables, but not all. Specifically bias related to respondent education level (college degree) continues to exist and needs additional scrutiny. State-to-state comparisons also show that potential bias differs by geographic locations of the samples.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Evaluating the Impact of Eliminating a Bounding Interview

Brett McBride *Bureau of Labor Statistics*

John Dixon *Bureau of Labor Statistics*

The Consumer Expenditure Interview Survey (CEQ) moved from a five-wave panel design to a four-wave design in 2015, eliminating the interview used to 'bound' a household's expenditure reports. Past research questioned the effectiveness of the bounding interview in limiting telescoping errors and projected that its elimination would result in significant cost savings. However, the elimination risked increasing the length of the first interview (which now included new questions previously asked in the bounding interview), and increasing later non-response rates due to a longer first interview. This poster investigates what effect, if any, the change had in reporting, interview length and response rates for the CEQ. In terms of reporting, the analysis compared April 2015 interviews that had a prior bounding interview with interviews in the same month that did not (due to how the new CEQ structure was implemented). It was hypothesized that the proportion of households reporting irregular and large expenditure categories would increase without the bounding interview. The second analysis focused on how the length of interviews administered in November 2014 changed one year later, after the consolidation of questions into four interviews. Response analysis compared response rates for households recruited for the first two interviews in 2014 with households recruited for the new (unbounded, longer) interview and the subsequent interview in the first half of 2015. The poster displays results showing the hypothesized increase in reporting for large expenditure categories. It charts increased year-over-year interview lengths for the (now) unbounded first interview. Finally, the poster displays the increased non-response in the second interview; noting this appeared less a result of changes in prior interview length and more as part of a secular trend in increasing non-response rates. This poster will be of interest to those concerned about telescoping errors and the impact of interview length.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Developing an Optimal Contact Strategy for the American Housing Survey

Aliza Kwiat *US Census Bureau*

Courtney Reiser *US Census Bureau*

It is an ongoing challenge to motivate people to respond to surveys. The American Housing Survey (AHS) plans to move from a Computer-Assisted Personal Interviewing (CAPI) data collection strategy to a mixed-mode approach. The planned mixed-mode approach will begin with a web self-response option, followed by CAPI for nonrespondents. AHS has recognized the importance of developing a strong contact strategy that will encourage people to utilize the web self-response option and turned to the Census Bureau to assist in developing an optimal contact strategy. In past cycles, AHS has relied upon a single mailing, which included a letter, FAQs, worksheet, and additional inserts. The challenge is to improve and update these materials so that recipients will notice, open, read and respond to the mailing. In this poster, we present our recommended strategy for optimizing AHS' contact strategy in the future. We will include results from qualitative research, including focus groups and cognitive interviews, which resulted in recommendations for interim improvements to the existing mailing materials for the 2017 CAPI-only production cycle as well as recommendations for new contact strategy materials to accompany the web response mode field test, and ultimately, AHS production. The research involved an extensive literature review of contact strategies, including the use of incentives and approaches towards encouraging record retrieval. We will also discuss the contact strategy approaches (i.e., types of contacts, number of contacts, and messaging) that we propose to be tested as part of the field test, with the end goal of developing a final strategy for implementation in AHS production.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Planning for the Future – Political Environments and Small Business Perspectives on Retirement Plan Sponsorship

Sarah Spell *The Pew Charitable Trusts*
Sarah Spell *The Pew Charitable Trusts*
Thomas Brassell *ICF International*
Andrew Blevins *The Pew Charitable Trusts*
James Dayton *ICF International*
Theron Guzoto *The Pew Charitable Trusts*
Brian Orleans *ICF International - Fairfax, VA*
Alison Shelton *The Pew Charitable Trusts*
John Scott *The Pew Charitable Trusts*
Robert Tortora *ICF International*

Social Security benefits were never meant to be the sole source of retirement income, but many Americans do not have sufficient private savings for a secure retirement. Though most save for retirement through workplace plans, as many as 30 million Americans lack access to such plans. Small businesses in particular find it difficult to afford or implement retirement plans. Policymakers are considering the feasibility of developing auto IRA plans, in which employees without access to an employer plan are automatically enrolled in an IRA plan managed by a third-party organization but sponsored by the government. However, the role of government is a politically divisive topic. The results of the 2016 presidential election offer the opportunity to evaluate the impact of the immediate political environment on support of third party sponsored retirement plans. Combining data from a nationally representative telephone survey of small, private-sector businesses with county-level voting results, we will assess whether there is a relationship between the immediate political environment and support of third party sponsored retirement plans. Specifically, we will use county-level voting results to identify the degree of political affiliation for the region (i.e., strong republican, lean republican, neutral, lean democrat, strong democrat). We will then assess the relationship between political affiliation and small business support of auto IRA plans overall and specifically when sponsored by the federal government, the state government, a mutual fund company, and an insurance company. We hypothesize that stronger political affiliations will have a significant relationship on small business support of third party sponsorship, but less of an impact at lower levels and when neutral. Understanding how political environment has any impact on small business attitudes will be of great value to policymakers evaluating the benefits and utility of proposing plans as well as the potential challenges of winning public support.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

NPS Or NOT?

Dinaz Jiwani *Boy Scouts of America*
Dan Warren *Boy Scouts of America*

Although the Net Promoter Scale (NPS) has been widely used in the business world to measure customer satisfaction and loyalty, it has some limitations when used in a greater evaluation model. The major drawback is with regards to binning responses into simplistic categories and limiting the analytical opportunities that provide variance, which is critical to consider while understanding trends and drawing actionable insights. The Boy Scouts of America are taking thoughtful steps away from traditional business customer experience metrics toward multi-level evaluation models which explore different segments including scouts, youth and parents beyond overall satisfaction and recommendation ratings to truly understand factors that drive satisfaction, loyalty and engagement. This approach affords the opportunity to introduce measures that are more multidimensional and allow for more nuanced interpretation. Also, in order to increase the usefulness of our data and the actionable steps we can take, we changed our yearly membership experience survey approach to a monthly data collection schedule. We accomplish this through rotational evaluation by exercising controls and using matched samples of our population to gain timely and frequent feedback. This revision creates a means to track metrics longitudinally and provide valid insights to stakeholders that drive improvement throughout the year. A comparison of these results with historic data that heavily relied on NPS suggested that a multitude of measures provides a holistic understanding of the customer journey which can assist in rationalizing the final reason for their satisfaction or loyalty outcomes.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

The Impact of Removing Instructions on Web Survey Responses

Rachel Horwitz *U.S. Census Bureau*
Elizabeth Nichols *US Census Bureau*
Julia Coombs *U.S. Census Bureau*
Casey Eggleston *U.S. Census Bureau*
Jessica Holzberg *U.S. Census Bureau*

Data quality is a large concern in all surveys, but in self-administered surveys in particular because there is not an interviewer present to answer questions or provide clarification. In addition to affecting data quality, this can also introduce mode effects in multi-mode surveys. Therefore, survey designers often include instructions for respondents in their self-administered modes. These instructions can be quite long, however. On paper forms, this is less of an issue, but with more surveys moving to the web and a higher proportion of web respondents using mobile devices, the screen real estate is coveted. Eye tracking research suggests that many respondents do not read the instructions provided. If this is true, it might not be necessary to take up valuable space on the screen to display them. This study offers a split sample of respondents a web survey with and without instructions and then compares the response distributions across treatment to determine what impact instructions have on survey estimates. Results will inform design decisions, particularly for mobile web surveys or surveys with a high percentage of mobile respondents.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Reducing Item Non-Response to Sexual Identity Questions Across Multiple Languages, New York City Community Health Survey

Michael Sanderson *NYC Dept of Health and Mental Hygiene*

Stephen Immerwahr *NYC Dept of Health and Mental Hygiene*

The New York City (NYC) Community Health Survey (CHS) is a telephone health surveillance survey conducted annually since 2002 by the NYC Department of Health and Mental Hygiene. CHS uses a fully overlapping dual frame design to obtain citywide and neighborhood-level health estimates, and is conducted in English, Spanish, Russian, and two Chinese languages (Mandarin and Cantonese, combined for this analysis). CHS questions on sexual identity have had modest levels of item nonresponse. In 2014, 6.1% of respondents had missing data for a question on sexual identity with three possible responses: straight/heterosexual, gay/lesbian/homosexual, or bisexual. Most persons with missing data answered "don't know" (4.2%), followed by "refused" (1.9%). Item nonresponse was significantly higher among respondents interviewed in Spanish (11.7%) or Chinese (14.3%), compared with those interviewed in English (4.3%) or Russian (5.3%). The weighted prevalence of sexual minority identity (not straight/heterosexual) was 4.0% overall, but <1% among respondents interviewed in languages other than English. Re-worded sexual identity questions were asked in 2015, following guidelines from the National Health Interview Survey (NHIS): using labels that respondents use to refer to themselves, avoiding labels some respondents do not understand, and follow-up questions to meaningfully categorize those answering "something else" or "don't know". Compared with 2014, item nonresponse in 2015 was lower overall (3.4%, a 44% reduction) and among respondents surveyed in English (2.1%), Spanish (9.2%), and Chinese (3.9%). With an added "something else" response category, the weighted prevalence of sexual minority identity was higher overall (5.8%, a 38% increase) and among respondents interviewed in English (6.6%), Spanish (4.5%), and Russian (2.8%). It remained <1% among those interviewed in Chinese. These results support NHIS recommendations and suggest both the need for caution in interpreting estimates of sexual identity among NYC adults (particularly among Asians) and an area for further analysis and improvement.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Business Payroll survey and Web Collection Strategy - Results of the transition from a paper based to a Web based self-response business survey and the use of paradata in shaping future collection strategy.

Anie Marcil *Statistics Canada*

Wade Kuseler *Statistics Canada*

At Statistics Canada, the implementation of a self-response Web Questionnaires as the primary collection mode has been a major priority for the organisation. To date more than 80 business questionnaires have transitioned from paper self-response to web self-response. The use of paradata from these surveys is very valuable and provides metrics to analyse and review the current collection strategies used. But how are we doing? Can we improve future collection strategies? The research efforts have focused on the impact associated with the effort being made to obtain response, reducing costs, maintaining high levels of quality and optimizing the use of the operational resources. This poster presentation looks at the collection results of Statistic Canada's Monthly Business Payroll survey and covers its Web Collection Strategy, provides examples of the results over time and outlines the future plans to test other collection approaches as we learn from experience.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

The effectiveness of a monetary incentive offer on survey response rates and response completeness in a longitudinal study

Shengchao Yu *New York City Department of Health and Mental Hygiene*
Howard EAlper *New York City Department of Health and Mental Hygiene*
Angela M Nguyen *New York City Department of Health and Mental Hygiene*
Robert M Brackbill *New York City Department of Health and Mental Hygiene*
Lennon Turner *New York City Department of Health and Mental Hygiene*
Deborah J Walker *New York City Department of Health and Mental Hygiene*
Carey B Maslow *New York City Department of Health and Mental Hygiene*
Kimberly C Zweig *New York City Department of Health and Mental Hygiene*

Achieving adequate response rates is an ongoing challenge for longitudinal studies. The World Trade Center Health Registry (WTCHR) is a longitudinal health study that periodically surveys a cohort of ~71,000 people exposed to the 9/11 terrorist attacks in New York City. Since Wave 1, the Registry has conducted three follow-up surveys (Waves 2-4) every 3-4 years and utilized various strategies to increase survey participation. A promised monetary incentive was offered to survey non-respondents for the first time in the recent Wave 4 survey, conducted 13-14 years after 9/11. We evaluated the effectiveness of incentive use five months after survey launch in improving the response rate without potentially damaging data quality. The study compared the likelihood of returning a survey for those who received an incentive offer and those who did not. Among those who returned surveys, we also examined whether those receiving an incentive notification had a higher rate of response completeness than those who did not. We found that a \$10 incentive offer was effective in increasing Wave 4 response rates. Specifically, the \$10 incentive offer was useful in converting the initially reluctant participants into respondents. The likelihood of returning a survey increased by 30% for those who received an incentive offer (AOR=1.3, 95% CI: 1.1, 1.4). Moreover, our results did not reveal any significant differences on response completeness between those who received an incentive offer and those who did not. In the face of the growing challenge of maintaining a high response rate for the WTCHR follow-up surveys, this study showed the value of offering a monetary incentive as an additional response rate enhancement strategy. Our findings also suggested that an incentive offer could be particularly useful at the near end of the data collection period when an immediate boost in response rate is needed.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Precinct Level Vote Data

Martin Barron *NORC at the University of Chicago*

Despite the general growth in easily accessible administrative data, precinct level vote data remains difficult to acquire and analyze. Election precincts (sometimes referred to as voting districts) are the smallest election jurisdiction in the United States and are the most geographically precise. Each U.S. address has a specific precinct associated with it and a distinct voting location (or polling place) associated with that precinct. Voting districts typically are relatively small, with one 2004 study finding an average of 1,100 registered voters per voting precinct. Because of the geographic precision of precinct results, a variety of analyses could take advantage of such data were it easily accessible. But it is not. There is no central repository for such data, no standardized format for it, and no mechanism for its collection. Instead precinct data is spread across numerous state and county websites (and for some states not available online at all), is stored in many different formats, and its data elements are not standardized. In this presentation we explore how a small team was able to cheaply and quickly collect, standardize, geographically map, and disseminate, 2012 and 2016 precinct-level presidential vote data. This paper discusses the “Extract-Transform-Load” (ETL) pipeline that allowed us perform these tasks with few resources. We provide a roadmap for projects acquiring scattered diverse data to produce useful analytic data efficiently and to widely share that data.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Hunting for auxiliary variables in the Census Planning Database Tract File

Clayton Knappenberger *Bureau of Labor Statistics*

Arcenis Rojas *Bureau of Labor Statistics*

Lucilla Tan *Bureau of Labor Statistics*

The motivation behind seeking higher response rates is to minimize the potential for nonresponse bias that stems from differences between survey respondents and non-respondents. However, higher response rates need not imply more representative responses unless the added respondents reduce the contrast between respondents and non-respondents. Measuring this requires information on characteristics on the complete sample (i.e. both respondents and non-respondents), which are difficult and expensive to collect. Information that is auxiliary to the primary purpose of the survey's data collection effort, and which are available on the complete sample, can be used to provide a partial measure of this contrast with Representativity (R) Indicators, which compare the composition of respondents to the complete sample with respect to a set of characteristics derived from the auxiliary information. In addition, auxiliary information that is predictive of survey response and the survey variables of interest are used in post-survey adjustment methods to reduce potential nonresponse bias. In the ongoing effort to better understand potential nonresponse bias in the Consumer Expenditure Interview Survey (CE), the CE is exploring the Census Planning Database (PDB) as a source of auxiliary information. The PDB comprises a range of housing, demographic, socioeconomic, and census operational data at the census block group and tract levels of geography. In this poster, we share our experience of learning to use the PDB Tract file with a household survey and the methods used to identify useful auxiliary variables from among the over 500 variables in the PDB Tract file. We also present preliminary findings on auxiliary information from the PDB associated with survey response propensity and total expenditures for the CE. Our lessons learned will be helpful to other survey practitioners interested in using the PDB.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Home Sweet Home: The effect of telecommuting on FEMA interviewer productivity

Kristin L. Brooks *Federal Emergency Management Agency*

Jessica Guillory *Federal Emergency Management Agency*

Brandi Lea *Federal Emergency Management Agency*

Emily Abbe *Federal Emergency Management Agency*

The ability to telecommute, or work from home, has become increasingly common in today's technologically-advanced work environment. The U.S. federal and state governments, for example, have encouraged this work arrangement through the introduction of legislation (e.g., Department of Transportation and Related Agencies Appropriations Act of 2000). Previous research supports telecommuting has favorable but modest effects on perceived autonomy, work-family conflict, and job satisfaction. When evaluating productivity, supervisors typically report an increase in job performance, although this may be confounded with only high performing employees receiving the privilege to telecommute. Additionally, a large proportion of these studies suffer from small sample sizes, have varying definitions of telecommuting, and often obtain conflicting results. The Federal Emergency Management Agency (FEMA) maintains a full-time staff of interviewers that administer various surveys to disaster survivors via Computer Assisted Telephone Interviewing (CATI). Most staff report to the office 1-2 times per week, and spend the rest of the week telecommuting. The current study examined whether 13 FEMA surveyors maintained similar productivity (e.g., survey completions, average interview length, number of attempts) while in the office vs. telecommuting. Paradata was used from May to November 2016, with all days of the regular workweek represented. Survey type and length were controlled for. In order to control for differences in individual performance, a series of dependent t-tests were performed for each productivity measure. Results showed that surveyors had significantly more completes per hour and shorter average interview length when telecommuting. Number of attempts per hour did not significantly differ between work environments. By incorporating objective measures of productivity and using a sample of interviewers with varied experience and skill, the current study helps clarify the practicality of telecommuting in a survey administration domain.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Speaking for Ourselves: A study engaging immigrant and refugee communities in applied research

Daniel Swanson *Wilder Research*

Nicole MartinRogers *Wilder Research*

Speaking for Ourselves is a community-based research project with and for immigrant and refugee communities in the Twin Cities that was spearheaded by Wilder Research and funded by our parent organization, the Amherst H. Wilder Foundation. We initiated this study with a kick-off event at which nearly 100 people who have an interest in immigrant communities attended and gave input on the topics to be studied as well as the methods of collecting data and reporting back to the communities. At this event, we recruited an Advisory Board of 20 individuals who are mostly from the cultural communities included in the study and who hold professional positions in a variety of sectors in the Twin Cities. Interviews were conducted in person and over the phone with over 450 Hmong, Karen, Latino, Liberian, and Somali immigrants and refugees by bilingual interviewers from the cultural communities that were included in the study. The sampling method was modeled after Respondent Driven Sampling, which was developed by public health researchers to find and survey “hard to reach” groups -- it relies on respondents’ own social networks to identify other potential respondents, and uses incentives to encourage participation and referrals. Survey questions asked about education, employment, health (including mental health), housing, transportation, safety, personal finances, resettlement, and social engagement to learn what the biggest needs, concerns, and strengths are of immigrant and refugee communities in the Twin Cities. We also invited community institutions to buy a question, which helped to ensure the results are used by critical stakeholders to make real changes that will improve access, improve programs, target outreach, debunk stereotypes, influence public policy, and otherwise improve the lives of immigrants and refugees in the Twin Cities. This poster will illustrate the pros and cons of this innovative and culturally responsive method.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Safety Screening in the NIS RDD Cell-Phone Sample

Becky Reimer *NORC at the University of Chicago*

Jacquelyn George *NORC at the University of Chicago*

Benjamin Skalland *NORC at the University of Chicago*

Becky Reimer *NORC at the University of Chicago*

Qiao Ma *NORC at the University of Chicago*

James A. Singleton *National Center for Immunization and Respiratory Diseases*

When dialing cell phones for CATI surveys, respondent safety is an important ethical consideration for survey researchers. Since the National Immunization Survey (NIS) started dialing cell phones in 2010, the cell-phone safety screener statement used on every cell-phone dial has been "If you are currently driving a car or doing anything that requires your full attention I need to call you back at a later time." Following guidelines published by the AAPOR Cell Phone Task Force in 2010, NORC's IRB suggested that the cell-phone safety screener be modified to avoid referring to specific activities such as driving and instead refer more generally to respondent safety. In addition, in 2015, the National Center for Health Statistics Ethics Review Board (ERB) requested that the safety screener be phrased as a question, requiring explicit agreement from the respondent rather than reading a statement and continuing as the default behavior. Because the screener is administered at the beginning of the survey, any change to the item wording could result in a change in respondent behavior, with subsequent impacts on response rates and costs. Therefore, before implementation of a new version, a large-scale experiment was conducted to test the original version of the cell safety screener against two alternate versions that incorporated the IRB and ERB feedback. We will describe the experimental design and results of the experiment, including the impact on survey response rates and how this work may apply to other CATI surveys that dial RDD cell-phone sample. As the cell-phone population continues to evolve, experimentation with cell-phone methodology is increasingly important.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Rounding Effect and the Use of Records in the CEQ

Taylor J Wilson *Bureau of Labor Statistics*

Safia Abdirizak *Bureau of Labor Statistics*

Based on the results from the 2015 Consumer Expenditure (CE) Field Staff Survey Analysis Report, 98.4% of Field Representatives stated that record use improves the accuracy of the interview. The primary purpose of this project is to conduct research regarding the use of records as it relates to the rounding effect of recall interviews. This paper tests two hypotheses, [1] the use of records reduces the rounding effect and, as a result, increases data accuracy and quality and [2] the rounded expenditure amounts and non-rounded expenditure are significantly different. The hypotheses are important in aiding large survey programs in understanding how respondent rounding will affect the underlying data quality of the survey responses. Currently, the use of records is not ubiquitous among respondents. Understanding the differences in consumer units (CU) that utilized records compared to those CUs that did not utilize records is critical to understanding the effectiveness of record use to improve data quality in the CE. We employ a digit screening method to isolate records which follow a pattern consistent with the literature on survey rounding. By identifying those records which have the highest probability of being a rounded record, we make an assumption that those records have actually been rounded. The findings will be of interest to survey methodologists and practitioners working in large scale survey operations with recall survey components and specific data quality goals.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Who Are We Most Likely to Reach With Digital Advertising?

Stephanie Baumgardner *US Census Bureau*

The Census Bureau first used paid advertising to increase awareness and motivate response leading up to and during the 2000 Census, which experienced the first ever increase in the mail response rate over the previous census. Because of the success in 2000, the 2010 Census also utilized paid advertising in print, radio, television, billboards, and online formats. During both the 2000 Census and 2010 Census, the American Community Survey (ACS) benefitted from the census advertising campaign with higher mail response rates. Outside of decennial years, the relatively small sample size of most surveys compared with the size of the general population make traditional broad-based advertising methods cost-prohibitive. With the advent of digital advertising strategies, however, there is now the opportunity to deliver promotional messages to individual households within a survey sample. The ACS offers a large enough national sample to field a test in early 2017 of such methods and determine whether they increase response rates. In preparation for this test, we use data from the 2016 ACS Content Test, a test of new and revised questions for the ACS conducted in the spring of 2016, to compare differences between households that can be linked to a digital profile and those that cannot. We will identify differences between households that can be linked in terms of their propensity to respond and use response data from the ACS to analyze select characteristics such as race/ethnicity composition, age, tenure, and educational attainment. Results of this research can inform the survey community on the characteristics of households most likely to be reached with digital advertising.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Correctly Cited Survey Data: The Unicorn of Public Release Research

Aimee Vella Ripley *Nielsen*

Sophie Bethune *American Psychological Association*

Allyssa Birth *Nielsen*

Correctly cited survey data in public release research is as elusive in the media as it is crucial. Data and survey results are everywhere today. They are a powerful tool in marketing, promotion, and advertising alike and proper citations give us critical information about the validity of the results being discussed. But properly cited data does more than simply tell us the who, when, and how of the interviewing process; it has the power to reveal the true sponsor of the research, discerns a properly researched topic from an improper one, and many other critical aspects when deciding whether to read a survey's results as truth. With a proliferation of content, it is becoming more and more common to see studies improperly cited and as researchers it can be difficult to catch and correct each individual instance. The problem further propagates as one story featuring improperly cited data spreads like a game of telephone. Correcting this issue has the power to increase the public's trust in and the influence of the research profession. In reaction to the challenges the industry is facing in our current landscape, Harris Poll (Nielsen) and the American Psychological Association created a simple, easy-to-follow guidelines for reporters and writers on how to interpret data for APA's Annual Stress in America survey, ask the right questions to fully understand a study's validity, and understand their role in combating improperly sourced data. The guidelines developed address grid questions, scaled questions, and multiple response questions. Specifically, what each type of question measures, what they do not measure, and how to properly report data from each type of question. During the presentation, we plan on outlining the common errors we encountered by reporters misinterpreting our data.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Using a Web Portal as a Hub for Respondents in Business Surveys

Aryn Hernandez *US Census Bureau*
Temika Holland *U.S. Census Bureau*
Michelle Karlsson *U.S. Census Bureau*

Currently, the U.S. Census Bureau sends separate URLs, user IDs, and passwords to business respondents for each economic survey request. In order to modernize and streamline the reporting process for business respondents, the Census Bureau has created the online Economic Respondent Portal. Over the next few years, all online economic surveys will migrate to this new portal, where respondents will only need to manage one set of login credentials of their own making when accessing any of the surveys they are responsible for completing. This is particularly useful to business respondents, who often are selected into multiple, repeated sample surveys. In addition, the portal provides respondents with a secure messaging center, survey help, and self-service options such as checking filing status, requesting due date extensions, delegating response tasks and data collection activities to colleagues, viewing reporting history, and more. These features may ease respondent burden by facilitating common business survey response processes, such as coordinating response timing with data availability and gathering data from multiple people and sources throughout the business. In this paper, we will discuss the Economic Respondent Portal's features, select results from two rounds of in-person usability testing, and how these results inform decisions about interface design and system functionality. In the first round of testing, the portal was presented as part of an overall online reporting experience and only account creation, survey linking, and delegation were explored. In the second round of testing, the portal was the sole focus of the interviews and all available features were tested asking users to think aloud while performing tasks, followed by more specific cognitive probing. In addition, we will offer some insights into how a Portal and its features may be adapted for Web surveys of households and individuals, particularly those requiring longitudinal panels where repeated participation is essential.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Evaluation of Online Listing Tool

James Cajka *RTI International*
James Rineer *RTI International*
William Wheaton *RTI International*
David Chrest *RTI International*

For some surveys frame generation using in field counting and listing is the gold standard. Because field enumerators physically visit the survey area and record each candidate location, the frame can be very accurate with a low degree of under or over coverage. The main drawbacks of in field enumeration however, are that it's expensive in terms of travel, accommodation, and personnel costs. Time and scheduling can also be an issue since the time to perform counting and listing can take weeks or months even if multiple crews are used. Safety of field staff is also a concern due to traffic, lack of familiarity with an area, and crime. RTI International developed an online listing tool that uses a combination of base maps, imagery, Google StreetView and 3D renderings to allow a user to capture locations and characteristics from their office. For our initial tool evaluation we selected 15 areas (segments) in central North Carolina to list commercial buildings by both in field and by use of the online tool. We compared the results in terms of coverage and time and accuracy. Although we encountered coverage gaps both between methods and between individual listers, the online tool performed very well overall. While there are many variables contributing to the coverage gaps, initial results indicate that online listing may perform better than in-person listing.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Implementation of 2014 Internet Test Results in the American Community Survey

R. Chase Sawyer *U.S. Census Bureau*

The American Community Survey (ACS) is continually testing ways to improve survey design and decrease respondent burden. In 2014, the Census Bureau conducted testing to determine if changes to the ACS Internet data collection instrument would increase response. Based on that research, the ACS implemented several changes in July 2016. The ACS now uses a security question instead of solely using a computer-generated Personal Identification Number to allow respondents to re-enter the instrument. In addition, the screens that transition between persons were improved to limit breakoffs. Finally, fill-in-the-blank response boxes were modified to increase response and lower the number of error screens presented to the respondent. This study determines if these adjustments have the same results that were observed in the 2014 Internet Test are replicated in a production environment.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

What Conditions Allowed the Trump Phenomenon

Jacob Nelson *Survey Sampling International*

Paul Johnson *Survey Sampling International*

Researchers commonly measure the importance of different policy issues among US voters with a rating scale or by ranking top issues. However, it is not common to measure satisfaction of current US direction on these policy issues separately. Instead researchers tend to ask satisfaction only on the general direction of the country. For example, the economy might be an important issue, but they could be satisfied with the current direction the economy is headed. This extra dimension creates a two by two grid that leads to different messaging in each quadrant. We hypothesize that the quadrant for these policy issues differed significantly in 2016 when compared to previous elections, giving Trump an opening to capitalize on. To demonstrate this difference, we will sample of voters through an online panel and split it into two groups. The first group will be assigned to recall the issues of the 2012 presidential election while the second group will be assigned the same issues in the 2016 presidential election. We will use two separate MaxDiff exercise to force differentiation in both importance and satisfaction of a set of policy issues. The resulting data will give us two-dimensional insights into the US political landscape, and explore how that landscape shifted between the two election cycles.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Evaluation of Algorithms to Determine Diabetes Type in Health Surveys

Sarah Lessem National Center for Health Statistics

National datasets are often used to provide population-level estimates of conditions such as diabetes and examine their comorbidities. There are different types of diabetes with different comorbidities, so it is important to understand who is included these estimates and analysis. Researchers generally use one of several algorithms to identify respondents most likely to have type 1 diabetes (T1DM) and assume all other respondents who report diabetes have type 2 diabetes (T2DM). In this article, I examine the components that go into algorithms, consider the reasoning behind each, and review potential problems with both the data quality and the rationale behind each algorithm components. I select seven potential algorithms and use NHIS, NHANES, and CHIS data to examine how the percentage of respondents identified as having T1DM changes by algorithm and agreement between each possible pair of algorithms. I focus on how individuals of different races are typed differently by each algorithm. I find tremendous variation in who is identified as having T1DM by each algorithm; with the most restrictive algorithms identifying 4.0% as having T1DM, and the least restrictive identifying 15.0%. Further, the agreement between algorithms varies substantially both by algorithm combination and by race. The algorithms generally have low agreement, with kappas $<.40$. Almost every combination of algorithms showed much poorer agreement for people of color, particularly Hispanics and Asians, than for Whites. In some instances, kappas indicate that two algorithms assigned the same diabetes type to people of color at no more than what would be found by chance levels. The results indicate very low consistency among algorithms in identifying diabetes type, particularly for people of color. As a result, depending on which algorithm is used, widely varying samples will be identified as having T2DM.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Examining Multimodal Transportation in El Paso Region: Results of the Regional Transportation Survey

Ipek Nese Sener *Texas A&M Transportation Institute*
Richard Lee *Texas A&M Transportation Institute*
Chris Simek *Texas A&M Transportation Institute*
Alfredo Sánchez *Texas A&M Transportation Institute*
Rafael Aldrete *Texas A&M Transportation Institute*

This paper summarizes a component of a multimodal transportation study conducted for the El Paso Metropolitan Planning Organization that included a regional population survey. The survey was implemented between September and October 2015, and the study area included all of El Paso County, Texas, and parts of Doña Ana and Otero Counties, New Mexico. The survey specifically focuses on examining the behaviors or motivations for using or not using the various modes of transportation, with a particular focus on alternative modes (i.e., walking, bicycling and public transportation). It is unique in that it is one of the first regional transportation surveys in the United States (to the best knowledge of the authors) contributing to the emerging area of health and transportation by integrating several elements of health and well-being. The results indicated the important role of understanding regional and population characteristics, as well as behavioral barriers and motivators, as an important first step in designing effective alternative travel programs and interventions. Several factors, including neighborhood environment variables, personal and social characteristics, and attitudes and habits (towards health and transportation) work together to influence residents' participation in alternative transportation. While the personal automobile is the current dominant travel mode in El Paso, residents see great value in extending public transportation infrastructure in the future. Improving connectivity, safety and the neighborhood environment were found to be particularly essential to encourage walking and bicycling in the region.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Examining the Effects of Accelerometry Device Type and Distribution Method in Two Consecutive Waves of a Longitudinal Study

Daniel Lawrence *NORC at the University of Chicago*

Lauren Bishop *NORC at the University of Chicago*

Kriston Koepp *NORC at the University of Chicago*

Katie O'Doherty *NORC at the University of Chicago*

Recent years have seen many population-based surveys take an interest in expanding their physical measures collection by asking respondents to wear a device that measures sleep and activity levels. However, because of the very diversity of devices and technologies that has made it increasingly feasible to include such measures, no uniform protocol has been developed for distributing these devices, achieving high response rates, and ensuring data quality. The National Social Life, Health, and Aging Project (NSHAP) has included a leave-behind measure of sleep and activity called accelerometry in two separate waves of data collection. Because each of the waves used a different accelerometer device and method of distribution, we can compare the advantages and disadvantages of each. This paper presents complete results for the accelerometry measure from both NSHAP waves in which such a measure was taken. We consider initial cooperation rates, final return rates, cooperation on other physical measures, amount of prompting required, respondent adherence to protocol, and respondent characteristics. Additionally, because in the most recent wave of data collection the accelerometry measure was offered both to panel respondents who had completed the measure in the past and to newly recruited respondents, we can examine the marginal effect of prior participation on respondent cooperation. The paper concludes with lessons learned about incorporating this type of physical measure into survey operations as well as methods for ensuring data quality.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Measuring the Effect of Immediate Post-Household Screener Completion Outcome on Survey Outcomes in a Population Study

Katie O'Doherty *NORC at the University of Chicago*

Daniel Lawrence *NORC at the University of Chicago*

Lauren Sedlak *NORC at the University of Chicago*

Anna Wiencrot *NORC at the University of Chicago*

In population-based surveys, it is sometimes tempting to conceptualize household screening as a completely separate effort from respondent recruitment and interview completion. However, if respondents identified as eligible upon completion of a household screener have the option to complete the interview immediately or schedule an appointment immediately, what impact does that post-screening outcome have on other study outcomes? For the recently concluded third wave of the National Social Life, Health, and Aging Project (NSHAP), NORC conducted a nationwide household screening effort to recruit a new cohort of respondents using a laptop that allowed for an interview to be launched immediately after completing the household screener. Because traditional household screening efforts often involve multiple visits, multiple devices, or other logistical challenges, there can be significant time or respondent burden between identifying a respondent as eligible and actually interviewing the respondent. This paper attempts to quantify the gains to be had if the time and respondent burden between screening and interviewing can be minimized by dividing this newly recruited NSHAP cohort into three screener outcome groups: respondents who completed the interview at the time of screening, respondents who scheduled an appointment at the time of screening, and respondents who neither interviewed nor scheduled an appointment at the time of screening. The paper presents comparative results between these groups in regards to completion rates, time to completion, cooperation rates for the physical measures in the NSHAP interview, respondent characteristics, and outcomes for eligible household members where someone else in the household completed the household screener. We also consider whether the relative difficulty of completing the household screener is an intervening variable, or whether it has an effect independent of post-screener outcome.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Day of the week adjustments for sample weights for NHANES 24-hour dietary recall

Te-Ching Chen *National Center for Health Statistics*

Joseph Goldman *USDA*

Jennifer Parker *National Center for Health Statistics*

The What We Eat in America is conducted as a partnership between the U.S. Department of Agriculture and the U.S. Department of Health and Human Services. Under the partnership, the National Center for Health Statistics is responsible for the survey sample design and all aspects of data collection. All National Health and Nutrition Examination survey (NHANES) participants are eligible for up to two 24-hour dietary recall interviews, one in the Mobile Examination Center (MEC) and another by telephone interview 3 to 10 days later. The NHANES participants are selected using a national probability design. To increase the number of participants for specific demographic groups, a multi-stage, unequal probability of selection design is implemented. Sample weights are constructed that encompass the unequal probabilities of selection, as well as adjustments for non-participation. Because food intake can vary by weekdays and weekends, special weights accounting for the day of the week of the food intake and component nonresponse for the 24-hour dietary recall are created for each 2-year data cycle. Beginning with NHANES 2011-2012, the NHANES sample design has included an oversample of Asian Americans. In part as more sampling domains, very small race-age-day of the week sample weight adjustment cells were found when creating dietary weights for NHANES 2011-2012 and 2013-2014. The purpose of this project is to compare alternative dietary weights for the 1st dietary recall based on different groupings of day of the week. Differences in dietary intakes between weekend and weekday interviews are assessed. While initial results confirmed known patterns of weekdays and weekend intakes, estimated differences between weekday and weekend intake using alternative weighting approaches are small. For example the difference average total energy intakes by weekends/weekdays among adolescent females ranged from 190 to 221 kcal across weighting methods, with p-values from t-tests ranging from 0.02 to 0.04.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Cognitive Testing for the Large Population-Based Survey About Eating Habits and Intestinal Illness

Matthew Thomas *ICF International*

Melissa Cidade *ICF International*

Emma Cogan *ICF International*

Naomi Freedner *ICF International*

ICF conducted a series of six cognitive interviews testing the navigation through and question wording of two versions of a large population-based survey about eating habits and intestinal illness. This survey aims to measure exposure to diarrheal vectors in the form of raw meats, vegetables, and others. This survey also seeks to identify unreported instances of diarrhea in the general population. We employed a framework that was dependent on evaluating survey questions against their measurement objectives, including what the question was attempting to measure, and to what level of accuracy respondents could provide data in response. This framework is dependent on a “think aloud” technique that encourages respondents to verbalize thoughts while engaged in a cognitive activity with little interjection by the interviewer other than to keep the respondent thinking out loud. In engaging respondents in this exercise, their cognition is slowed while not impacting their task performance. We found that respondents struggled with a number of items. Most prominently, respondents were likely to focus too closely on examples provided in food consumption questions while disregarding other food products that might be applicable to a given question. When asked to refer back 7 or 30 days, respondents often added or dropped days to the time frame. Additionally, proxy parent respondents often showed little knowledge about their child’s food consumption habits when they were not in their presence.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Predictors of survey completion time among adolescent respondents using mobile and desktop devices

Matthew Thomas *ICF International*

Rachel Schmidt *ICF International*

Olivia Saucier *ICF International*

James Dayton *ICF International*

The widespread availability and popularity of mobile devices has provided an additional platform for administering web surveys. It is unclear how differences between mobile and desktop devices impact survey completion time among youth. Our implementation of a longitudinal web survey of youth provides the opportunity to assess differences in survey completion between mobile and desktop users. On behalf of a large university, ICF administered a series of 4 web surveys among approximately 2,600 students in grades 6 through 10. These surveys were formatted and optimized to be completed on both desktop and mobile devices. We conducted a paradata analysis using data from one wave in order to better understand survey completion times of students participating in the study. In particular, we investigated the effects of device type and demographic variables on completion times. Preliminary results suggest that students who access the survey on a desktop device were more likely to complete the survey in a shorter amount of time than those who completed on a mobile device. There were several key predictors of survey completion time, including, age, ethnicity, and GPA. Older students, as well as students who reported receiving mainly A's in the previous semester tended to finish the survey in less time than their peers.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Are Urban Areas Always Less Religious?

Becka Alper *Pew Research Center*

Jessica Hamar Martinez *Pew Research Center*

The 2016 presidential election demonstrated the well-documented differences between urban and rural areas spanning a range of demographic characteristics and social and political attitudes. And while we also know that people living in urban areas tend to be less religious than those who live in rural areas, little is known about how religious beliefs and practices differ across urban areas. Using extensive data from the Pew Research Center's 2014 Religious Landscape Study, which measures religiosity and public opinion among more than 35,000 U.S. adults, we examine how religious affiliation and beliefs and practices vary between rural and urban areas, while controlling for other important characteristics such as gender, education, race/ethnicity and political partisanship. This analysis will allow us to assess the extent to which urban and rural areas vary in terms of religious affiliation and beliefs and practices, yielding important information about how they may intertwine with demographic, social and political attitudes. And considering how important geographic areas are in sampling, this analysis also provides important insight into a key, though often overlooked, way that urban areas can vary.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Impact of Observed Neighborhood Characteristics on Interview Quality

Sara Walsh *NORC at the Univ of Chicago - Chicago, IL*

Jennifer Satorius *NORC at the University of Chicago*

Lauren Bishop *NORC at the University of Chicago*

Anna Wienrot *NORC at the University of Chicago*

Interviewer observations on a respondent's home environment, such as housing unit conditions, structural features of housing units, and features of the neighborhood, such as safety, evidence of crime, or presence of trash have been increasingly used to examine response and nonresponse rates of households (West, Kreuter 2013). Groves and Couper (1988) indicate that physical impediments such as an intercom or security features may obstruct contact with the respondent and thus cooperation, whereas type and state of housing unit as part of the socioeconomic environment have been associated with cooperation, with well-maintained dwellings and neighborhoods having lower refusal rates than multi-unit housing structures. While much research has been done investigating the impact of observed neighborhood characteristics on survey response, little research has gone on to examine the impact of neighborhood characteristics on interview quality upon survey completion. For the National Social Life, Health and Aging Project (NSHAP), a study of aging and health from a multidisciplinary perspective, interviewers are required to complete an interviewer comments questionnaire immediately (or as close to immediately as possible) from conclusion of a completed case interview. Among other variables, interviewers are asked to record observational data about the respondent's general neighborhood and dwelling unit. Since the interviewer comments questionnaire is only available upon survey completion we have the unique opportunity to focus our analysis on interview quality, such as within questionnaire completion rates, cooperation with biospecimen and physical measures collection, and task completion such as cognitive measure exercises. Comparing interview observation of neighborhood characteristics against interview quality will allow us to gain insights that will help to improve interviewer training, impact staffing plans, and design field approaches for specific areas. As overall survey costs are rising, finding improvements that will reduce costs and improve efficiencies, especially for in-person studies, is of increasing importance to our field.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Supplemental Letters and Response Rates: The Effects of Mode and Length

Lena Le *Social and Economic Sciences Research Center*

Matthew Strawn *Social and Economic Sciences Research Center*

Thom Allen *Social and Economic Sciences Research Center*

A visitor study was conducted in Southeast Colorado to determine the economic impact of visitors to heritage tourism sites in the region. Participants were recruited onsite at predetermined locations throughout the area using intercept sampling approach based on methodologies developed over the course of the Visitor Services Project (VSP) for the National Park Service. Unlike VSP surveys, however, this study used a push-to-web approach with participants randomly assigned to one of four experimental groups. All participants received a postcard reminder and all non-respondents received a paper version of the questionnaire. A supplemental letter was sent in conjunction with the postcard and questionnaire mailings. The mode of dissemination and length of the supplemental letter was manipulated in a 2x2 factorial design. Participants received either a long or a short version of the letter either by postal mail or via email. This experiment attempts to determine how the variations in an additional contact affect response rates and data quality in survey research.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Impact of Advance Letters on Response Rates and Data Quality in a Statewide Dual-Frame Survey

Eva Aizpurua *University of Northern Iowa*

Mitchell Avery *University of Northern Iowa*

Ki Park *University of Northern Iowa*

Rod Muilenburg *University of Northern Iowa*

Jill Wittrock *University of Northern Iowa*

Mary Losch *University of Northern Iowa*

Many studies have illustrated the effectiveness of pre-notification letters in increasing response rates to telephone surveys. However, most studies assessing the impact of advance letters have been carried out using landlines, and the evidence in dual frame samples is scarce. This is important because the proportion of cell numbers used in dual frame telephone surveys continues to rise yet little is known of the impact that advance letters have on cell phone respondents. In this study, we assess the impact of an advance letter on response rates in a statewide dual-frame survey on perceptions and experiences with healthcare. In a large dual-frame telephone survey of Iowa residents, a sub-sample of those with addresses (both for landline and cell) were randomly assigned to receive an advance letter (n=800). The other half (n=800) did not receive any pre-notice. Response and cooperation rates will be compared between the two modes, as well as between those participants in the treatment group who recalled having received the letter, and those who did not. In addition, we will analyze the effect of advance letters on various aspects of data quality, paying special attention to measures of item non-response. The effect of advance letters on the number of call attempts required for interview completion will also be presented. Implications of the results for the use of advance letters in dual-frame telephone surveys will be discussed.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Correcting for Co-Viewing in the Digital Space: Providing a More Accurate Measurement

Elaine Perrin *Nielsen*

Lindsey Rabham *Nielsen*

Matt Vanlandeghem *Nielsen*

Big data provides a myopic perspective on consumers without the use of panels. However, it often reflects machines instead of the persons behind those machines; a panel designed specifically to measure human behavior, not only its server footprint, is needed to complement big data. Moreover, consumer viewing of video content is increasingly moving towards the digital space. Measurement companies can easily match demographics of viewers by using cookies from data enrichment providers (DEP). However, DEPs cannot determine whether there are multiple people looking at the screen simultaneously. They only capture the single demographic associated with the DEP cookie. This occurrence of co-viewing isn't limited to television content, it also occurs in any online activity, such as advertising. Nielsen, a leading global measurement company, measures desktop co-viewing within its demographically-balanced computer-metered panel, where individual household members are required to log into the meter whenever they use the computer. This panel allows Nielsen to estimate DEP cookie gaps based on what is observed in the panel and what can't be seen on a non-metered device, including co-viewing. Nielsen has developed a methodology that calculates a "Co-Viewing factor" to be applied on DEP cookie data to adjust for unique multiple viewers. For example, a fraction of video views by a female aged 35-39 will have a child viewing with her. That child's video view should also be counted, but is not detected by a DEP cookie. Nielsen's new methodology is able to capture the child's viewing. This methodology involves taking the estimated proportion of co-viewers viewing the screen, based on how often people from different demographics view simultaneously, and adding it to the total number of page views/sessions by demographic bucket. Including co-viewing allows for a more accurate measurement of who is viewing content on desktops.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Perceptions of Intimate Partner Violence in Europe: Assessing Individual and Country-level Factors

David Vazquez *University of Castilla La Mancha*

Eva Aizpurua *Center for Social & Behavioral Research, University of Northern Iowa*

Jennifer Copp *Florida State University*

Jorge Javier Ricarte *University of Castilla La Mancha*

Gender-based violence remains a major social and public health problem in the European context, and intimate partner violence is the most common form of violence against women. Public awareness of the prevalence of intimate partner violence against women in society is an important first step to trigger public responses, and inform policy and programmatic efforts. Social and personal accountability for violence against women stems from public perceptions of the problem as widespread, and as posing a threat to the wellbeing of women in society. That is, lacking public concern, the issue seems unimportant and action unnecessary. Factors explaining public perceptions of violence against women are multiple and can be identified at different levels, including individual- and country-levels. An appropriate understanding of variation, both between and within countries, calls for a multilevel approach. This study analyzes individual and country-level factors influencing women's perceptions of gender-based violence. Multilevel modelling was used to study 42,000 women, nested in 28 member states of the European Union (EU). The outcome variable assessed women's perceptions of the prevalence of intimate partner violence in their countries. Individual-level predictor variables included victimization and knowledge of awareness-raising campaigns, as well as personal and sociodemographic characteristics. At the country-level, we accounted for the absence/existence of specific laws or political initiatives, an index of gender equality, and the prevalence of violence against women. The findings underscore the complexity of violence against women, and the need for future research that further establishes the influence of a broad range of factors across multiple levels of aggregation, including the potential for cross-level effects. Future work should also analyze public perceptions of violence against women by assessing public views of the types of violence that are most common.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Data Editing in Large-Scale Surveys

Elise Comperchio *NORC at the University of Chicago*

Caitlin Finan *NORC at the University of Chicago*

Megan Stead *NORC at the University of Chicago*

Chris McCormick *Centers for Medicare & Medicaid Services*

Shannon Corcoran *Centers for Medicare & Medicaid Services*

The Medicare Current Beneficiary Survey (MCBS) is a continuous, multipurpose survey of a nationally representative sample of the Medicare population, conducted by the Centers for Medicare & Medicaid Services (CMS) through a contract with NORC at the University of Chicago. This poster illustrates the process of reviewing and editing complex longitudinal survey data for the MCBS, which includes approximately two thousand variables covering a variety of topics, such as health status and functioning, health care use and expenditures, and health insurance coverage. The poster identifies the various sources of erroneous data, including interviewer errors, questionnaire programming errors, and issues in data processing. It covers methods for conducting two critical phases of longitudinal data review: database structure check and variable logic check. The poster then describes the implementation and testing of global and case-specific edits to address errors in survey data. In addition, the poster presents innovative methods for streamlining and automating quality control of data review and editing. Such methods include the development of 1) shared macro libraries to facilitate and standardize data review tasks, 2) a centralized master program to verify editing, 3) rule reports showing the number of cases edited, and 4) comparisons of each round of editing to verify fixes and identify any unintended changes. Lastly, future directions in innovation to further increase efficiency and effectiveness in cleaning complex survey data are discussed.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

The impact of multitasking on survey data quality: observations from a statewide telephone survey

Jill Wittrock *University of Northern Iowa*

Erin Heiden *University of Northern Iowa*

Eva Aizpurua Gonzalez *University of Northern Iowa*

Ki Park *University of Northern Iowa*

Mary Losch *Ctr for Social & Behavioral Res Univ of Northern Iowa*

The proliferation of mobile phones and hands-free technology has introduced new avenues for respondent distraction during a telephone interview. More frequent and intensive multitasking and greater environmental distractions are two of several possible ways in which survey data quality can be affected. In both instances, the respondent is answering questions while engaged in something or with someone else. We expand on prior research on distractions and multitasking during telephone interviews in multiple ways. First, we examine the extent to which respondents with a particular set of characteristics (e.g., younger, children at home) are more or less prone to multitasking. Second, we distinguish among different types of multitasking activities (e.g., technical, manual) and whether there is an impact on survey data quality based on the type of distraction. Our dataset comes from a statewide dual-frame telephone survey of adult Iowans who were queried on topics related to awareness, perceptions, and attitudes toward STEM education. At the end of the interview, respondents were asked what other activities, if any, they had been engaged in while responding. About half (55%) of respondents reported engaging in some type of multitasking activity while completing the interview. The analysis will examine the extent to which distraction type impacts tendencies to give “don’t know” or “not sure” responses and interview length. Implications for survey practice will also be discussed.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Evaluating Behavioral Weighting Controls for Online Convenience Panels

William DeShong *Nielsen*

There are broadly two ways of measuring digital audience: server-based and panel-based measurement. Server-based measurement excels at volumetrics for content publishers willing to implement measurement software, but falls short on demographic information. Panel-based measurement provides an in-depth view of demographics and captures all viewing behavior of panelists. But it suffers from all the typical challenges associated with sampling methodology. Specifically, convenience panels have an inherent bias generated from both representation and recruitment methods. Nielsen, a leading data and measurement company, leverages a large online panel and performs weighting techniques to correct for sampling biases. Nielsen's online panel consists of both a high quality probability panel and convenience panel of US internet users that are used to ensure both depth in measurement and behavioral bias correction. A multi-step weighting procedure has been implemented that corrects for demographics and behavioral bias. Behavioral weighting controls are selected that help reduce the inherent bias. The universe corresponding to each control is set from the high quality probability panel. An emphasis will be put on how these controls were selected and tested.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

The Shifting Composition of Party Association During the 2016 Presidential Election

Hyesun Choung *University of Wisconsin-Madison*

Josephine Lukito *University of Wisconsin-Madison*

Trevor L Kniaz *University of Wisconsin-Madison*

Hyungjin Gill *University of Wisconsin-Madison*

Douglas McLeod *University of Wisconsin-Madison*

Song Wang *University of Wisconsin-Madison*

The purpose of this study is identifying the distinct sub-groups within each of the two major parties in the U.S. The 2016 presidential election has seen the two parties fissure into many subgroups such as democratic socialists, the Alt Right movement, evangelical conservatives, and the Never-Trump movement. These rifts distinguish politicians from their voter base, to the degree that politicians may not have endorsed the nominees for their respective parties. As a result, voters within a party have become increasingly divided over political issues and candidates. These divisions may be influenced by factors such as worldview, demographics, and personal values, ultimately influencing the ever-evolving organization of the parties themselves. The heightened rejection of the political status quo (from which the left-right political spectrum is devised), a re-evaluation of where partisans fall on the “political spectrum” is necessary. Given the contentiousness of the discourse within parties, one thing has become clear: there are distinct factions within both parties that render the traditional labels of Democrat and Republican into gross generalizations. Such generalizations obfuscate the significance of significant differences of different groups that fall under those party umbrellas. Based on recent scholarship, we identify voter clusters by analyzing data collected from the national survey (N = 2,670) and try to see to what extent their characteristics align with our current understanding of partisan voters. The clusters are finally linked to patterns of media use, issue orientations, and voting behavior.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Removing bottlenecks and helping interviewers succeed; our journey based on lessons from "The Goal: a process of ongoing improvement"

Howard Speizer *RTI International*

Societal and technology changes make it increasingly more difficult to locate sample members, engage interest, and secure survey participation. For interviewer-administered surveys, these challenges directly impact the interviewer; their duties, performance, and morale. We are motivated to help the interviewer with intelligent, fit-for-purpose, survey designs, training and support but more needs to be done. This research uses operations management strategies based on the theory of constraints, and introduced in the popular business text *The Goal* (Goldratt, 1984) , to systematically identify and remove process bottlenecks that hinder performance. We use a model of the survey data collection process to identify and focus on eight key attributes that impact the interviewer's success (abilities, knowledge, commitment, engagement, attention, priorities, review, and support). Evaluating the steps in the process for each attribute uncovers bottlenecks and other areas for improvement. Targeted initiatives and associated outcome measures based on the model and process review are designed and implemented. We use expert review to rank the expected impact of these improvements on desired outcomes. We also report initial results from the improvement initiatives and lessons learned from the process.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

The Effect of Telephone Survey Center Staffing on Productivity and Survey Results.

Enrijeta Shino *University of Florida*

Enrijeta Shino *University of Florida*

Christopher McCarty *University of Florida*

Mark Girson *University of Florida*

The main objective of this study is to examine the effect of telephone survey center staffing schedules on data quality and productivity. Some centers continue to staff based on characteristics of landline samples. Under this model the majority of calls are scheduled from 6 pm to 9 pm on weekdays and on weekends, while the weekday morning and afternoons are much more limited. Some prior work has been done on identifying scheduling procedures that reduce data collection costs and nonresponse bias using landline random digit dialing (RDD) household surveys. The main finding is that the high contact and completion rate for the first call are on weekday evenings. However, in the era of cellphones is it necessary to make the highest number of calls in the evening? With the introduction of the dual-frame and cellular RDD samples, and the large increase in popularity of cell phones, we posit that survey lab staffing should be reassessed. If we are contacting different types of respondents in the morning versus evening, we should be aware of these differences and evaluate their effect on the data quality. The change in quality of the data across shifts is measured as the difference across shifts between the respondent demographics. We address whether the bias issue could be solved using post-stratification weights. Productivity is measured as the number of dials per complete interview. To make this study's results applicable to more than one sample method, we analyze data collected from three different studies using cellular RDD, dual frame RDD, and listed samples. To test whether contact time affects our understanding of societal issues, using the Florida Consumer Sentiment Index survey data from 1985 to 2015, we compare a series of economy, health, and state government evaluation models controlling for demographics and interview time.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Exposure to counter-attitudinal TV news and affective polarization: Moderating effects of emotional stability and need for cognition

Yanqin Lu *Indiana University*

Drawing on the 2012-2013 American National Election Studies panel data, this study explores the effect of counter-attitudinal TV news use on affective polarization, which refers to partisans' tendency of viewing out-party members negatively and in-party fellows positively (Iyengar & Westwood, 2015). The results reveal that exposure to counter-attitudinal news programs reduces an individual's level of affective polarization, suggesting the important role of cross-cutting exposure in promoting a healthy democracy. Further, such effects are found particularly stronger among individuals with a higher level of need for cognition (NFC) and lower level of emotional stability. Theoretically, these findings provide support for the notion that partisan news use influences affective polarization through two possible mechanisms: attitude rehearsal and affective learning (Garrett et al., 2014). Specifically, counter-attitudinal news attenuates the level of affective polarization through extensive thinking among those with a higher level of NFC; individuals with a lower level of emotional stability tend to become less polarized by learning negative emotions toward their in-party politicians from counter-attitudinal TV news.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Extending the Spiral: Inter and Intrapersonal Effects of Sharing One's Opinion in a Hostile Opinion Climate

Emily Van Duyn *The University of Texas at Austin*

Spiral of silence theory has long served as a model to explain the stifling of opinions over time. Within this canon there exists mixed evidence for the theory's two mechanisms, fear of isolation and perceived disagreement, with some studies offering support (Neuwirth, 2000; Scheufele et al., 2001; Yun & Park, 2011) and others arguing they serve as weak predictors of expressing one's opinion (Huang, 2005; Petric & Pinter, 2002). As a result, this tradition has largely considered internal processes before expression rather than the effects after speaking out. This study looks to explore the inter- and intrapersonal outcomes after expressing one's opinion, particularly effects on fear of isolation and willingness to speak out in the future. Through an experiment simulating conditions for spiral of silence to occur, this study extends the theory by testing the influence of interpersonal feedback on emotion and future willingness to speak out. Drawing from symbolic interactionism and the looking-glass self, this study hypothesizes that positive interpersonal feedback will relate to positive emotion, decreased fear of isolation, and increased willingness to speak out in the future. Findings from this relationship hold implications for how spiral of silence processes may relate to interaction, in addition to avoidance, to the extent that interpersonal feedback motivates or discourages individuals from speaking out in the future.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

The Impact of Partisanship on Beliefs about Global Warming: The Mediating Roles of Perceptions of Scientific Consensus, News Media Use, and Trust in Scientists

Soohee Kim Stanford University

Prior research suggests that American public is skeptical about the existence of global warming despite the overwhelming consensus among scientists about the existence of global warming. While many researchers have noted partisan identity as a critical factor driving people's beliefs of global warming, little is known about how and why partisan identity influences individuals' beliefs of global warming. In this study, we examine the potential mechanism underlying the effects of partisan identification, particularly focusing on the interaction between partisanship and other critical constructs – such as perceptions of scientific consensus, trust in scientists, and partisan media use. Using data from a nationally representative sample of U.S. adults, this study finds that partisan identity both directly and indirectly influences beliefs of global warming. In particular, the finding shows that partisanship has a considerable indirect effect on individuals' beliefs of global warming via all three constructs – perceptions of scientific consensus, trust in scientists, and partisan media use, with perceptions of scientific consensus the most influential factor. Additional analyses demonstrate that partisanship also moderates the impact of partisan media use: the negative effect of FoxNews watch on perceptions of scientific consensus was greater among Republicans than among Democrats, thus having a more negative influence on Republican's belief of global warming. Policy and research implications of these findings are discussed.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Understanding Public Attitudes toward Immigration Policy: Political Ideology, Self-Interest, and Political Knowledge

Tianshu Zhao *University of Illinois at Chicago*

Timothy Johnson *University of Illinois - Chicago*

The issue of immigration continues to trigger variations in people's sentiments. Public opinion appears complex concerning attitudes toward undocumented immigrants. While many studies have discussed the impacts of self-interest, political ideology, political knowledge, and other determinants on public attitudes, there is little research trying to categorize and specify the roles of self-interest and political ideology. In this study, we examine two distinct measures of objective and subjective self-interests, create measures of operational and symbolic ideologies, and explore the main effects of political knowledge toward immigration attitudes. Through running SEM models based on hypotheses, we find that objective and subjective self-interest show distinctive correlations with immigration attitudes, so do operational and symbolic ideology. These findings suggest that objective self-interest and operational ideology predict immigration attitudes better than subjective self-interest and symbolic political ideology do. Also using SEM models helps us to better understand the complicated relationships among these measures. In doing so, this study extends current models and traces a theoretical and empirical path bridging the literature in political science and public opinion concerned with immigration.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

The Politics of Queer Religion

Royal Gene Cravens *University of Tennessee*

In the United States, the debate between religion and sexuality often juxtaposes Christian conservatism, and its stand for traditional morality and 'family values' against modern sexual liberation in what is often characterized as a "culture war" (Fiorina, Abrams, and Pope 2005). Despite this perception, many LGBT people, whether Jewish (Schnoor 2006), Muslim (Rahman 2010; Minwalla, Rosser, Feldman, and Varga 2005), or Christian (McQueeney 2009; Rodriguez and Ouellette 2000), have succeeded in 'integrating' their sexual and religious identities. Through socialization, religious belief, and the lack of it, affects political attitudes on a variety of subjects, as well as participation in the political process (Schwadel 2005; Campbell 2004; Verba, Scholzman, and Brady 1995; Wald 1987). Few quantitative studies exist, however, that examine the effects of religious identity on sexual minority politics. That is, sexuality and religious identity can conflict. Given such a situation, does religious conflict influence political participation in LGBT political activities? Overall, religious affiliation and participation are thought to work as mobilizing structures which have a positive effect on political participation (Brady, Verba, and Scholzman 1995) while specific denominational effects appear to be the most significant in predicting political ideology (Conger 2009). Given the rejection of homosexuality by many denominations, are the positive effects of religiosity on political participation mitigated when belief conflicts with identity? Using data from two nationally-representative surveys of self-identified LGBT people and informed by resource mobilization theory, the identity threat model of stigmatization, and an intersectional approach, evidence suggests that religiosity is associated with increased political participation among LGBT people; however, participation in LGBT-specific political activity is less likely among those who experience conflict between their religion and sexuality. This finding is reinforced by denominational effects which suggest LGBT Protestants are significantly less likely to participate in LGBT political activity.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Watchdogs & Lapdogs: Donald Trump's Taming of the News Media

William R. Davie *University of Louisiana*

The 2016 presidential race generated as much heat as light, with Americans' interest reaching its highest point in any presidential campaign since 1988. Pew Center data showed 85% of the American voters gave a lot of thought to the election, up from 71% in 2012. Yet voters were more frustrated than ever before, even though the economy was recovering, employment was up, military intervention was down, and health care services were more accessible. Two sources of voter rage were politicians and the press. In a campaign routinely punctuated by personal attacks, 68% of the electorate claimed that insulting an opponent was not fair game. President-elect Trump, however, used this technique to generate more press coverage than would be expected. This content analysis shows how two types of news media -- watchdogs and lapdogs -- framed the race to Mr. Trump's advantage. The view from the United States' rival in economic affairs, China, shows how this played into that country's world view.

May 19th, 2017
3:15 PM - 4:15 PM
Concurrent Session P

Poster Session #2

Survey Breakoff in Mixed-Device Web Surveys

Anke Metzler *Darmstadt University of Technology*

In recent years the use of smart phones to complete Web surveys has grown steadily. Thus, survey researchers are facing new challenges. Among others, previous research has shown that breakoff rates are higher in mobile Web surveys than in PC Web surveys. Breakoff rates can be minimized by using an adaptive design approach that accommodates all devices. However, even Web surveys with an adaptive design approach indicate higher breakoff rates among respondents using a smart phone than among respondents using a PC (Mavletova & Couper, 2013; Buskirik & Andrus, 2014). Several factors that cause breakoff in PC Web surveys such as sociodemographic characteristics and attitudes about surveys are also responsible for breakoff in mobile Web surveys (Antoun, 2015). However, the focus of this poster is on causes that are unique to breakoff in mobile Web surveys. I assume that due to device specific characteristics survey design features such as timing of the invitation, number of reminders and questionnaire design features have a different influence on survey participation behavior in mobile Web surveys compared to PC Web surveys. Furthermore, the respondents' device use such as context, multitasking, costs, screen size and mobile Web use (frequency and range of activities) also affects survey participation behavior in mobile Web surveys. Respondents of two Web surveys (n=4,620; n=944) conducted among university applicants were asked whether they were willing to participate in a follow-up survey. In each Web survey more than 75 percent agreed. Two experimental follow-up studies will be conducted and respondents will be either invited to a mobile Web survey or a PC Web survey to assess the reasons behind the higher breakoff rates in mobile Web surveys compared to PC Web surveys. Results will indicate specific factors determining cooperation in mobile Web surveys.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Maintaining Data Quality for International CAPI Surveys (Panel)

Maintaining data quality for international CAPI surveys

Martin Wulfe *MWulfe Consulting*

Beth-Ellen Pennell

Martin Wulfe *MWulfe Consulting*

Beth-Ellen Pennell *Survey Research Center, University of Michigan*

Zeina Mneimneh *Survey Research Center, University of Michigan*

Yu-chieh (Jay) Lin *Survey Research Center, University of Michigan*

Gina-Qian Cheung *Survey Research Center, University of Michigan*

Clark Letterman *RTI International - RTP, NC*

Safaa Amer *RTI International - RTP, NC*

Jamie Cajka *RTI International - RTP, NC*

Aimee Benson *University of North Carolina, Carolina Population Center, Chapel Hill, NC*

Meghann Jones *Ipsos Public Affairs*

Mark Andrews *Ipsos Public Affairs*

Michael Wild *World Bank*

Misha Lokshin *World Bank Group*

Galina Zapryanova *Gallup*

Johanna Godoy *Gallup*

International CAPI surveys present a unique set of challenges and opportunities for the maintenance of data quality. In addition to those challenges for PAPI surveys, such as language, culture, and logistics, they also include voice and data communications, device compatibilities, and others. These surveys also represent opportunities not available in other modes, such as immediate responsiveness to poor quality and access to paradata.

Collecting Rich Paradata to Monitor Data Collection Quality in Challenging Contexts

Beth-Ellen Pennell

This presentation will describe innovative uses of rich paradata to monitor production and quality indicators in household surveys in developing and transitional countries. Technology is increasingly facilitating new approaches to more efficient production and better quality monitoring through the collection and monitoring of rich paradata (process data). This diffusion of technology not only allows for immediate access to the survey and process data (including call records) to monitor field work quality but has also facilitated the use of other applications. These applications include self-administered modes (e.g., audio computer-assisted self-interview [ACASI]), digital audio recordings, global positioning systems (GPS) for collecting contextual information or live tracking of interviewer travel to households, areal photography for sample selection, and the collection of various anthropometric data using digital devices, among other examples. With these innovations come new challenges, however. The presentation will discuss advantages, challenges and lessons learned across a diverse set of projects as well as make recommendations for new developments in this space. The presentation will provide examples from projects in five very different settings: China, Ghana, India, Kingdom of Saudi Arabia and Nepal.

Using Geo-Sampling to Improve Data Quality in Household Surveys

Clark Letterman, Jamie Cajka

Since 2014, RTI has used geo-sampling - a new technique for sampling households - in over 70,000 CAPI interviews in low- and middle-income countries. Geo-sampling improves data quality by leveraging the speed and accuracy of geographic information systems (GIS). In this presentation, we describe how geo-sampling can improve the quality and representativeness of household samples. The geo-sampling approach uses GIS data to select 1km x 1km primary grid units (PGUs) and smaller secondary grid units (SGUs) within each PGU. Interviewers then conduct a census of eligible households within the SGUs. Interviewers use digital maps on a tablet to navigate to the SGUs and establish its boundaries, while passive GPS recordings allow the management team to monitor where interviews take place. Under the geo-sampling approach, interviewers conduct a census of all eligible households in sampled clusters. The quality of the sample is thus dependent on the interviewers' ability to follow rules, including (1) identifying whether they are in the correct cluster, (2) establishing whether a given building is eligible or not, and (3) following within cluster fieldwork procedures, such as sampling in multi-story buildings. Our presentation focuses on the training and quality monitoring necessary to ensure that interviewers follow these rules, especially when working with interviewers and organizations accustomed to using random walk or household lists for household sampling.

Transparency enhancement to improve management and quality control of face to face interviewing using Ipsos' custom platform

Meghann Jones

Computer assisted personal interviews (CAPI) has finally, after much anticipated wait, become the standard for face to face (F2F) data collection in less developed countries where paper surveys were the standard just 5 years ago. This advent has, without doubt, improved the quality of F2F data immensely by reducing interview errors. Ipsos, as a global organization with offices in 87 countries, has developed a custom CAPI system, and has completed over 100 studies in 27 pilot countries by November 2016. It is planned to implement the system across all Ipsos offices by the end of 2017. The impetus for the development of this system included the desire to centralize the programming and data capture systems, to reduce the management and data processing burden on multi-country or multi-contractor surveys, create greater accountability of field staff, and to provide management tools to field offices. A central component of this strategy was to develop a system that maximized transparency by providing a project portal that allows managers to view metadata, survey results, and quality metrics in real time whether in-country or in an office across the globe. This paper will present an overview of the functionality of this custom system and present mini-case studies of how the tool is being used. This paper will also formulate lessons learned from our pilot implementation and outline our continual improvement effort.

Adventures in mode change: Upgrading from PAPI to CAPI

Aimee Benson

Several factors are contributing to the accelerated use of CAPI over PAPI in the developing world: availability of cheaper tablets with longer battery life, availability of multiple software options for data capture, and the demand for data in an ever-faster timeframe. As organizations make this shift, they experience both the benefits (e.g., enhanced data quality; faster data availability; greater interviewer satisfaction) and the challenges (e.g., technological glitches; increased need for upfront planning; limited

in-country capacity) that accompany it. In October 2016, the USAID-funded MEASURE Evaluation project launched a CAPI data collection effort as part of an impact evaluation of the USAID-funded Twiyubake program, which operates in 12 districts of Rwanda. 3,068 households currently enrolled in Twiyubake and 1,580 households from control areas were randomly selected for inclusion in the survey. Seventy enumerators and ten field supervisors were trained on how to use a menu-driven CSPro 6.3 data capture system to administer four interconnected survey instruments using Dell Venue Pro 8 Windows tablets. An external lookup file enforced case integrity for case management. A separate menu-driven CSPro system was installed on larger Windows tablets used by the field supervisors. The supervisors' system included case assignments and a suite of CSPro batch programs to check the completeness and quality of each questionnaire. This menu-driven system was used on the supervisors' tablets to create encrypted archives and to transfer the data to UNC's SFTP server. During training, several technical and logistical challenges arose and were addressed by a diverse team of programmers and IT specialists. This presentation will describe some of those challenges and how they were resolved.

Facilitating adaptive spatial cluster sampling through CAPI and real-time Monitoring: Experiences from a survey on informal businesses in Harare, Zimbabwe

Michael Wild

Rare and hard to reach populations, like informal sector enterprises, pastoralist households or homeless people are difficult to survey as they are not included in standard sampling frames and their location is usually not known a priori. Additionally these types of populations are often strongly clustered. To overcome the first problem regarding the frame, we will construct a spatial sampling frame (grid), which will allow to conduct probability surveys even without having a comprehensive and current standard population frame available. To address the second characteristic of a clustered distribution of this type of population without using a very large sample size, we will apply a special type of sampling design, namely stratified adaptive cluster sampling. In adaptive cluster sampling, a starting cluster (a grid cell) is randomly selected, and if within the starting cluster the number of survey units exceeds a predefined threshold, all other clusters around this starting cluster are surveyed as well. This approach was rarely applied so far in face-to-face establishment surveys due to challenges in the field management. However, we expect to be able to address these challenges by using a well programmed CAPI questionnaire, implemented with the World Bank's Survey Solution software package, the use of a navigation software, which includes the delimitations for each grid cell as well as by real-time field monitoring.

Interviewer Characteristics and Social Desirability Bias in Face-to-Face Interviews

Galina Zapryanova

Do interviewer demographics such as age, gender and education affect respondent answers on culturally sensitive questions? Interviewer characteristics and their effects on response rates and response quality have long been the subject of inquiry in public opinion research. We present new data from nationally representative surveys in 15 countries in Latin America (N=1000 per country). During the 2016 wave of the Gallup World Poll, additional data on interviewer demographics was collected for selected regions. This paper presents the results from Latin America – we examine whether interviewer gender, age and level of education affect respondent answers on substantive topics such as employment, well-being, security and political views. Additionally, we examine whether respondent demographics condition the likelihood of susceptibility to interviewer effects. Social desirability theory holds that respondents are more likely to give answers that would be viewed favorably by others – this makes the presence of interviewers in face-to-face interviews particularly likely to generate such effects.

However, not all topics are susceptible to social desirability bias and country differences often exist even within the same region when it comes to what is considered socially or culturally desirable. By comparing across countries and across substantive topics within country, we are able to test the presence or absence of interviewer effects by taking into account both individual-level characteristics and the country context. With face-to-face interviews still being the dominant mode of data collection in many parts of the world, it is particularly important to examine the methodological consequences of interviewer and interview characteristics. This paper will thus contribute to our understanding of social desirability bias in the Latin American context and the methodological challenges it translates into during survey fieldwork.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Redirected Inbound Call Sampling (ICS) - A New Survey Research Tool (Panel)

Redirected Inbound Call Sampling (RICS) - A New Survey Research Tool

Karol Krotki *RTI International*
Scott Richards

RICS is based on the possibility of intercepting incorrectly-dialed calls and replacing the curt termination message with an invitation to complete a survey. The methodology is nonprobabilistic and open to bias but on the other hand it is extremely inexpensive and quick. The number of such calls that can be intercepted on a daily basis in the USA numbers in the millions. This panel consists of four presentations: the first outlines the methodology and describes potential for future development. The next three presentations describe the experience of three survey organizations in using this methodology to implement national surveys – most importantly, they focus on bias issues. The fourth presentation discusses RICS as an efficient means of screening for rare and hard-to-reach population as well as a tool for bio-surveillance.

Redirected Inbound Call Sampling (RICS) – Detailed Methodology

Scott Richards

RICS is a new nonprobability sampling methodology that recruits individuals that make a call that did not reach its intended recipient to participate in a survey. Over 5 billion calls a month placed in the U.S. and Canada do not reach their intended recipient. We refer to these calls by the acronym MIDI—misdialed, incomplete, disconnected, inbound calls. Reconnect Research has a nationwide network of telecommunications companies that redirect their MIDI calls to the Reconnect Research IVR (Interactive Voice Response) platform. Callers hear an intercept message such as “Please take our national health survey. Your call couldn’t be completed and was redirected to this survey”. In addition to IVR, inbound callers can be surveyed by a live interviewer or sent to a web site to complete a web based instrument. The speaker, CEO of Reconnect Research, provides details on how MIDI calls become survey respondents, the operational details involved in using ICS to implement surveys, the scalability of the system, and what future developments might include.

Evaluating Bias in a Survey Using Redirected Inbound Call Sampling (RICS)

Burton Levine, Karol Krotki

In RICS, telephone calls that fail to connect to their intended target are routed to a data collection system. The allure of RICS methodology is in the low cost and speed. By using an interactive voice response system, thousands of interviews can be collected on the first day of data collection with a cost less than one-tenth of a traditional outbound telephone interview. RICS produces a nonprobability sample and potentially contains selection bias as well as the other sources of error found in probability surveys. To quantify bias, we conducted an RICS survey with 7 questions that duplicate demographic questions asked in the 2015 American Community Survey (ACS) and 15 health outcome questions that were asked in the 2015 National Health Interview Survey (NHIS). We quantify bias by treating the population estimates from the ACS and the NHIS as correct and quantify the deviation from the unweighted RICS results for demographic characteristics and the weighted RICS results for the health

outcomes. We also examine bias in a multivariate analysis. Because RICS is less expensive than outbound telephone surveys, for the same budget, more interviews can be collected with RICS methodology than outbound telephone surveys. We describe the budget where RICS methodology will produce estimates with lower mean squared error than an outbound telephone survey using our estimate of bias, making various cost assumptions and assuming that an outbound telephone survey has zero bias.

Measuring Public Opinion with Redirected Inbound Call Sampling (RICS)

Courtney Kennedy, Kiley McGeeney, Nicholas Hatley

RICS has been proposed as a low-cost alternative sample design for conducting surveys. An early test of this methodology, which routed such calls to an interactive voice response (IVR) system, showed some promising results with respect to the demographic profile of the unweighted sample and the comparability of health estimates relative to those collected in CDC's Behavioral Risk Factor Surveillance System (BRFSS) (Levine, Krotki and Bobashev 2016). If RICS can, in fact, produce data that are on par with much more expensive random digit dial (RDD) designs, this methodology stands to have a dramatic impact on survey research. Much more research is needed, however, to evaluate the properties of estimates from these inbound calling designs. Key research questions include: How does respondent attention and engagement in such surveys compare to CATI surveys with RDD samples? Do adults reached through this methodology differ in important ways from the general population or are they more akin to a random sample? Is this methodology susceptible to the same over-representation of civically engaged adults that has been documented with other survey designs? This presentation reports on a RICS study to address such questions. We used this methodology to conduct three national surveys of adults, each with n=1,500 respondents. The surveys featured a number of common public opinion questions, supporting comparisons of public opinion estimates from this new technology to estimates from dual frame RDD CATI surveys. The study also featured a number of benchmark questions from high quality, high response rate federal surveys, supporting estimates of bias. Implications for public opinion researchers in particular are discussed.

Redirected Inbound Call Sampling (RICS) – Pilot Test Results and Caller Reactions

Sarah Dipko, Eric Jodts

This paper presents results of testing a new non-probability sample source of inbound calls, in essence a "river sample" of callers. Inbound callers interested in completing the survey can be connected with IVR surveys or to telephone interviewers. A 14.5 percent response rate (AAPOR RR3) was reported for the earliest test of RICS as a sample source for research purposes, with a survey conducted using IVR (Levine, Krotki and Bobashev, AAPOR 2016). The paper describes a ten-day RICS test conducted in July 2016. Live interviewers were used to conduct the study in order to better understand the circumstances and reactions of the callers in the RICS sample. What types of calls were being made when callers reached the survey request? How was the survey request and experience (for respondents) perceived by the callers? Assumed and observed rates are presented, and characteristics of the sample are compared to national estimates for demographics, geographic representation, and health measures. Cost and management issues are addressed such as out of scope calls (repeat calls from identical phone numbers or those from outside the United States) and significant fluctuations in the volume of the inbound call flow to interviewers. The survey experience for inbound callers is examined using survey data, paradata, and case notes combined with insight from an interviewer debriefing held after the study concluded.

Redirected Inbound Call Sampling (RICS) and Rapid Surveillance - Questionnaire Design and IRB Issues

Georgiy Bobashev

RICS allows one to administer rapid short surveys by redirecting misdials, incomplete and disconnected calls to an interactive voice response (IVR) system. Data produced by RICS can be collected quickly and cost-effectively, but before it becomes commonly accepted it needs to be evaluated in terms of representativeness and sensitivity to questionnaire design. We conducted several evaluation studies and present the results of calibrating RICS surveys to the US Census and sequential comparisons to (1) another RICS to check consistency, (2) a probability-based phone survey Behavioral Risk Factors Surveillance System (BRFSS), (3) household survey National Household Survey of Drug Use and Health (NSDUH), and (4) Non-traditional community-based surveys that address emerging issues. We show that while the results of the comparisons are promising, more rigorous research is needed to address potential biases, some of which are related to the timing of survey and the way questions are asked. We discuss the issues related to questionnaire design, sensitive topics, informed consent, and the protection of human subjects.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

At Least a Dozen Things We Learned from the 2016 Elections (Panel)

At Least a Dozen Things We Learned from the 2016 Elections

J. Ann Selzer *Selzer & Company*

Anthony Salvanto *CBS News*

Doug Schwartz *Quinnipiac University Poll*

Mark Blumenthal *SurveyMonkey*

Four election polling experts talk about the challenges (and rewards?) of polling in the presidential election that challenged conventional wisdom at every turn. The panel includes: J. Ann Selzer, president of Selzer & Company and pollster for The Des Moines Register, and Bloomberg Politics, among others, Anthony Salvanto, Ph.D. is director of elections for CBS News. He heads the Decision Desk, Doug Schwartz, director of the Quinnipiac University Poll, and Mark Blumenthal, head of election surveys for Survey Monkey.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Recent Experience with Adaptive Design in Federal Surveys (Panel)

Recent Experience with Adaptive Design in Federal Surveys

Peter V Miller *U.S. Census Bureau*

Peter V. Miller

Stephanie Coffey *U.S. Census Bureau / JPSM*

Jim Dahlhamer *National Ctr for Hlth Statistics*

Benjamin Reist *US Census Bureau*

Andy Peytchev *University of Michigan*

Evolution of Adaptive Survey Design in Federal Surveys Adaptive survey design employs frame data, paradata and accumulating response data to make course corrections in data collection procedures to achieve particular data quality and/or cost goals. This approach differs from traditional data collection protocols, first, in that it aims beyond simply maximizing the response rate and/or staying within the survey budget. Second, adaptive design involves continuous survey management -- monitoring the characteristics of interviewed cases and allocating collection effort strategically to the remaining cases. These efforts seek to improve the quality of the achieved sample during data collection, maintain response rates and constrain costs. This panel begins with an overview paper that traces development of adaptive design in surveys conducted by the Census Bureau. The presentations then examine recent experience with adaptive survey design in three federal surveys: the National Health Interview Survey (NHIS), the Survey of Income and Program Participation (SIPP) and the National Survey of College Graduates (NSCG). Application of adaptive design in all three surveys seeks to improve sample balance during data collection and maintain response rates within a fixed cost context. The NHIS and SIPP are large CAPI interview collections. The presentations on these surveys involve experimental results from experiments on "case prioritization" -- the practice of allocating more interviewer effort to cases that, if completed, will improve the achieved sample balance, and giving less effort to cases that will contribute to sample imbalance. The NSCG is accomplished through Web, mail or telephone interviews. The paper on this survey presents experimental results of mode switching, between self-administered (Web, mail) and interviewer administered (telephone) interviews, in order to achieve better sample balance. The experimental findings from these studies suggest that application of adaptive design can achieve the goals of improved data quality within fixed cost constraints.

Growth of Adaptive Survey Design at the U.S. Census Bureau

Peter V. Miller

The past five years have seen noteworthy growth in research on adaptive survey design at the Census Bureau. Beginning with the development of necessary data resources (characteristics linked to frame data and different types of paradata), studies have progressed to randomized trials of adaptive data collection approaches, compared to traditional fieldwork. Adaptive approaches studied have included substituting administrative data for interviews, prioritizing interviewer workloads and using different survey modes to focus data collection efforts. Promising findings point to the adoption of adaptive survey design in some surveys in the near future. This paper reviews the progress of research across a number of surveys and introduces the work presented in other papers in this panel.

Adaptive Design in the National Survey of College Graduates: Findings from the 2015 Experiment and Prospects for 2017

Stephanie Coffey

Falling response rates are driving surveys to seek novel contact strategies (or combinations of contact strategies) to slow the decrease in response, or at least to reduce the cost spent on unproductive cases. The National Survey of College Graduates has implemented dynamic adaptive design experiments in the last two rounds, 2013 and 2015, and will once again be running at test in the 2017 round of data collection. The 2013 test was only 4,000 cases and primarily operational, to test whether the Census Bureau had the capability to pull together disparate data collection systems to react to near-real-time survey paradata and assign cases to data collection treatments that were “off path” from the standard data collection schedule. Success in 2013 led to an expanded test in 2015 where fewer, larger interventions were made on nearly 20,000 cases in order to determine whether these changes could significantly shift data quality and operational metrics, including response rate, cost, and representativeness. This talk presents results of the 2015 experiment, including the estimated effect of various interventions on the previously mentioned evaluation metrics. The results of these analyses were incorporated into the 2017 design as a way to automate decisions on whether adaptive design interventions should be executed. The 2015 results demonstrate that the experimental interventions applied to the treatment resulted in significant improvements in representativeness, while either increasing or maintaining overall response rate. We will present results on improvements in sample balance, and interventions’ effects on data collection costs, as well. Our results demonstrate that adaptive design can have significant effects on markers of data quality and cost.

Implementing Adaptive Design in the National Health Interview Survey: A Case Prioritization Experiment

Jim Dahlhamer

Large federal surveys have experienced a steady drop in participation over the past two decades. Confronted with declining response rates and ever-tightening budgets, survey organizations are turning to responsive or adaptive survey designs to reduce cost, minimize nonresponse bias and/or variance, and maximize response rates. In this paper, we describe an adaptive design experiment conducted in the third calendar quarter of the 2016 National Health Interview Survey (NHIS). Through case prioritization, the primary goal of the experiment was to improve sample representativeness (or minimize nonresponse bias) within existing cost constraints, while maintaining current response rates. We address several research questions pertinent to the experiment. First, did interviewers alter their behavior consistent with the priorities assigned to their cases? Were more contact attempts made on high priority versus low priority cases? How did the effort invested in high and low priority cases compare to the effort expended on control cases (to be worked “business as usual”)? To what extent did treatment interviewers, relative to control interviewers, vary the days and times of contact attempts on high priority cases? Did they move their attempts to more productive time slots (e.g., evening hours and weekends)? Second, what impact did the experiment have on response and completion rates? Were response rates held steady or improved upon compared to prior months of 2016? How did response and completion rates vary between the control and experimental groups? And third, was greater sample representativeness (as measured by the R-indicator) achieved among experimental compared to control cases? We conclude with a discussion of results and their implications for administering case prioritization with large face-to-face interview surveys.

Using Adaptive Design to Prioritize Cases in the Survey of Income and Program Participation

Allison Zotti

In recent years, decreasing response rates for computer-assisted personal interviewing (CAPI) have forced many organizations to study new survey methodologies. Adaptive design methodologies have emerged as a framework for tailoring contact strategies to cases, such as case prioritization. The Survey of Income and Program Participation (SIPP) conducted an experiment that prioritized cases for the field interviewers on their workloads. This aim of this experiment was to increase data quality, in a fixed cost environment, by prioritizing cases less likely to be located, cases with less administrative record information, and cases less represented in the respondent population. SIPP is a nationally representative longitudinal household survey that gathers information on income and program participation. Each survey panel is interviewed over a multi-year period. Because it is a longitudinal survey, SIPP may struggle with panel attrition, leading to unbalanced response data. This experiment includes two consecutive interview periods of SIPP, conducted in 2016 and 2017. Each interview period includes 1,200 field interviewers and 35,000 cases. Field interviewers conduct all SIPP interviews, either in person or by phone. Interviews can be quite lengthy, estimated to take up to one hour per adult household member. During the first half of data collection, static business rules are used to prioritize cases that were less likely to be located or had less administrative record information. During the second half of data collection, a composite score is used to assign high and low priorities. The composite score is comprised of each case's propensity to respond and its balancing propensity, which were determined using propensity models and R-indicators, respectively. Comparisons of response rates and R-indicators were observed throughout data collection to determine the effects of the case prioritization experiment. Metrics were also developed to determine the field interviewer's compliance with the case prioritization guidelines.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Marijuana and Public Opinion Change (Panel) Organized by PAPOR

Marijuana and Public Opinion Change

Floyd Ciruli *Ciruli Associates*

Floyd Ciruli

Zac Auter *Gallup*

John Cohen *SurveyMonkey*

Stuart Elway *Elway Research*

Lunna Lopes *Public Policy Institute of California*

Recreational marijuana is on the move around the country. It was approved in four states in 2016 by mostly narrow votes and now is legal in states with more than 60 million people, or about 20 percent of the country. Pollsters will describe the shift in opinion favoring legalization, some of the future opportunities and road blocks it may face, and status of public opinion in states that approved it in 2012.

After Legalization, It's Time to Change the Question

Floyd Ciruli

A substantial majority of Colorado voters remain steadfast in their support for the legalization of recreational marijuana. But there are numerous signs of stress and public resistance to its spread across communities and through the commercial process of manufacture to sale. Polling to capture the stress and in communities resisting its spread recommends different questions from those developed pre-legalization.

Evolution of Opinion About Marijuana Legalization in the Northwest

Stuart Elway

After being among the first states to legalize medical marijuana (1998), Washington began circulating initiative petitions to legalize recreational marijuana in 1999. It took until 2012 for the measure to pass. Neighboring Oregon followed in 2014. The presentation will look at the evolution of Northwest opinion about marijuana legalization.

Legalize it! Examining the Predictors of Support for Marijuana Legalization in California

Lunna Lopes

In 1996 California became the first state to legalize medical marijuana through Proposition 215. An attempt to legalize the recreational use of marijuana failed at the ballot box in 2010. Yet in the 2016, advocates for legalization emerged victorious as Californians passed Proposition 64. Using data collected through the Public Policy of California's Statewide Survey series, this paper examines the changing attitudes among Californians toward marijuana legalization, comparing support for legalization in 2010 and 2016. This paper explore changes in attitudes towards marijuana among different subgroups including age, education, income, and race/ethnicity. We then use logistic regression analyses of datasets from 2010 and 2016 to explore whether the predictors of support for legalization have changed.

Trends in U.S. Marijuana Attitudes and Use, 1969-2016

Zac Auter and Jeffrey M. Jones

Since the late 1960s, Americans have become increasingly likely to support legalizing marijuana use, with a solid majority now in favor of it. This change reflects greater support among more recent generations than among older generations, but also increased support within each generation over time. In addition to the support of its legalization, there has been a dramatic increase since the 1960s in the percentage of Americans who report having ever smoked marijuana, and an increase in the last few years in the percentage who say they use the drug.

Which States are Next to Legalize Marijuana - 50 State Survey

Sarah Cho

We will examine which states would be ripe to pass a legalization initiative with national/50-state data on the issue from SurveyMonkey.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Using Interaction Coding to Understand and Improve the Survey Measurement Process (Panel)

Using Interaction Coding to Understand and Improve the Survey Measurement Process

Nora Cate Schaeffer *Univ of Wisconsin Survey Ctr*
Kristen Olson
Kristen Olson *University of Nebraska - Lincoln*
Jolene Smyth *Univ. Nebraska - Lincoln*
Antje Kirchner *RTI International*
Allyson Holbrook *University of Illinois - Chicago*
Timothy Johnson *University of Illinois - Chicago*
Sharon Shavitt *University of Illinois at Urbana-Champaign*
Young Ik Cho *University of Wisconsin -- Milwaukee*
Noel Chavez *University of Illinois at Chicago*
Saul Weiner *University of Illinois at Chicago*
Isabel Anadon *University of Wisconsin--Madison*
Jennifer Dykema *Univ of Wisconsin Survey Ctr*
Dana Garbarski *Loyola University Chicago*
Nora Cate Schaeffer *Univ. of Wisconsin, Survey Ctr.*
Ian Wall *University of Wisconsin--Madison*
Dorothy Edwards *University of Wisconsin--Madison*
Polly Phipps *Bureau of Labor Statistics*
Robin Kaplan *Bureau of Labor Statistics*
Brandon Kopp *Bureau of Labor Statistics*

Although survey research has witnessed a surge in the use of self-administered web and mail surveys over the past decades, interviewer-administered surveys continue to be an important part of the data collection landscape, particularly when measurement is demanding. We study interaction in the survey interview to uncover the problems interviewers and respondents encounter in performing their tasks, how they attempt to surmount those obstacles, and whether, when, and which actions affect the data. This panel features presentations that explore current issues and highlight recent findings using interviewer-respondent interaction coding (often referred to as “behavior coding”) to understand and improve the survey measurement process. Findings from the presentations provide immediate and practical recommendations for writing and evaluating survey questions and training interviewers.

The Effect of Question Characteristics, Respondents and Interviewers on Question Reading Time and Question Reading Behaviors in CATI Surveys

Kristen Olson

Asking questions exactly as worded is a fundamental part of a survey interviewer’s role. By reading a question fluently and exactly as written, the interviewer allows the question to be asked as intended by the researcher. Reading questions exact may decrease the risk of measurement error on survey questions and contribute to rapport (Dykema, et al. 1997; Garbarski, et al. 2016). Thus, understanding the risk of measurement error requires understanding how different interviewers, respondents, and

types of questions may trigger question reading problems. In this paper, we evaluate the joint effects of question, respondent and interviewer characteristics on question administration time and question asking behaviors. First, we examine how traditional question features affect the time to ask survey questions and how interviewers vary in their administration of these questions. Second, we investigate how choices in visual design features that require interviewer decisions contribute to accurate question readings by interviewers. Finally, we examine whether the effects of these features vary over interviewers and respondents. To examine these questions, we use the Work and Leisure Today (n=450, AAPOR RR3=6.3%) survey and the Gallup US/Japan Public Opinion survey (n=467, AAPOR RR1=7.4%). Using behavior coded data, we parse the question reading time from the rest of the question-answering time, as well as examine question reading behavior in particular. We use cross-classified random effects models that account for the nesting of question times within questions, respondents, and interviewers. In both studies, question characteristics account for roughly 20% of the variation in question asking behaviors and around 66% of the variation in question asking time. Question characteristics such as type and length of question, and question reading level predict question-asking time, whereas characteristics related to task complexity and visual design significantly affect exact question reading. We conclude with implications for interviewer training and questionnaire design.

Respondent Behavior and Survey Satisficing

Allyson L. Holbrook

Previous research has used respondent behavior in a variety of ways to study the survey response process. Research on survey satisficing suggests that when respondents are unmotivated or unable, they may engage in specific behaviors (e.g., selecting no-opinion or don't know options when explicitly offered) in order to give satisfactory answer to the survey questions without carefully and completely going through all the cognitive steps involved in answering survey questions. Researchers have also used other behaviors identified using behavior coding (whereby respondent verbal behaviors are coded from transcripts or recordings) to identify respondent problems with answering survey questions. A respondent who asks for clarification may be having difficulty with understanding the question and a respondent who gives a response that does not meet the question objectives may be having difficulty with mapping their response on to the response format given. In this paper, we examine the relationship between these two measures of the survey response process. Specifically, we use data from a survey in which equal numbers of non-Hispanic Whites, non-Hispanic Blacks, Mexican-Americans and Korean-Americans were interviewed face-to-face in the laboratory. Respondents also completed PAPI instruments both before and after the interview. Interviews were audio and video recorded and respondents' verbal behaviors were coded. During the face-to-face interview, measures of survey satisficing (including acquiescence response bias, no-opinion responding, and nondifferentiation) were obtained. We examine whether survey satisficing is associated with respondent behaviors coded using behavior coding (e.g., are respondents who show acquiescence response bias more likely to demonstrate specific behaviors identified using behavior coding?). These findings have implications for interviewer training because interviewers could be trained to identify behaviors that predict survey satisficing and be trained to encourage respondents to think carefully in order to avoid the reduction in survey data quality that is a consequence of survey satisficing.

Do Interviewer and Respondent Behaviors Predict Measurement Equivalence: Comparing Measurement of Trust across Racial/Ethnic Groups

Isabel Anadon

Features of the survey measurement process may affect responses from respondents in different racial, ethnic, or cultural groups incongruously. For example, in their analysis of the response process for a set of common health-related questions, Holbrook et al. (2006) reported greater comprehension difficulties among African-, Mexican-, and Puerto Rican-Americans than Whites. When responses from multiethnic populations are combined, such variability in responding could bias results or increase variable error. The purpose of the current study is to examine the survey response process among black, Latino, American Indian, and white respondents answering a set of questions designed to measure trust in medical researchers. Preliminary analysis indicates that a twelve-item measure of trust lacks measurement equivalence (i.e., questions do not appear to measure the concept in the same manner across the subgroups of respondents defined by their race/ethnicity). Although lack of measurement equivalence is a common problem, we know of no studies that have examined patterns of interaction between the interviewer and respondent in order to determine whether scale items with more measurement error are associated with greater interactional problem indicators. In the current study we (1) code for behaviors that have been demonstrated to be associated with measurement error (e.g., respondents displaying comprehension or mapping difficulties); (2) test for differences among subgroups in the frequency with which behaviors are displayed; and (3) describe whether the behaviors vary by the questions associated with measurement invariance. Data are provided by the 2013-2014 “Voices Heard” survey, a computer-assisted telephone survey designed to measure respondents’ perceptions of barriers and facilitators to participating in medical research. Interviews (n=410) were conducted with a quota sample of respondents nearly equally distributed into across the following subgroups: white, black, Latino, and American Indian. Interviews are transcribed and coded for the presence of potentially problematic interactional behaviors.

Exploring Interviewer and Respondent Interactions Surrounding Sleep Questions in the American Time Use Survey

Polly Phipps

The American Time Use Survey (ATUS) collects data on how people spend their time. Interviewers use a set of scripted, open-ended questions to walk respondents chronologically through their activities, including their start and stop times, who they were with, and where each activity took place, over the prior 24-hour period. The complex and detailed nature of this task is challenging for many respondents, and often requires interviewers and respondents to resolve uncertainties through verbal exchanges. One activity recorded in the ATUS has received a great deal of interest lately – sleep duration. Diary reports of sleep tend to exceed those of other national surveys that use stylized questions to collect data on how many hours per night people sleep. We focus on possible sources of measurement error in the time diary, including difficulties arising during the response process, such as respondent’s ability to recall or estimate time, in tandem with interviewer direction and assistance surrounding the reporting and recording of sleep activities. We systematically coded 104 ATUS interview transcripts to gain an understanding of issues related to interviewer and respondent behaviors that may affect the ATUS sleep estimates. We code numerous features of the diary interview, including questions, probes, answers, respondent recall strategies, uncertainty surrounding time estimates, as well as unscripted conversation that may contribute to error in the ATUS. Our findings suggest that the interactions, especially those surrounding when a person falls asleep, are often complex. This indicates that respondents may have difficulty recalling or estimating reports of their sleep. In addition, we found that interviewers often use unscripted probes that may inflate ATUS sleep estimates. We outline concepts and tasks that respondents and interviewers struggle with, practices that could be associated with sleep measurement issues, and offer possible solutions to reduce error.

Going off Script: How Interviewer Behavior Affects Respondent Behaviors in Telephone Surveys

Antje Kirchner

Despite attempts to fully standardize survey interviewing, interactions between interviewers and respondents are typically characterized by deviations from the ideal 'question-answer-neutral acknowledgement' sequence (Schaeffer and Maynard, 1996). Interviewers often engage in suboptimal behavior, such as misreading a survey question or engaging in directive probing. Respondents might refuse to answer a question, ask for clarification, or offer irrelevant information. Both can contribute to a breakdown in the response process. Several studies have investigated these behaviors in isolation, but not how they unfold over an interaction or how interviewer (mis)behaviors may trigger respondent (mis)behaviors. This paper investigates whether interviewer behaviors that facilitate efficient progress through the interview, such as reading the question exactly as worded, lead to improved response behavior. We expect that the effect of interviewer behavior on respondent behavior is moderated by question characteristics such as question length, reading level, unknown terminology, or sensitivity. To examine these hypotheses, we use the Work and Leisure Today Survey (n=450, AAPOR RR3=6.3%) and the Gallup U.S./Japan Public Opinion Poll (n=467, AAPOR RR1 = 7.4%), both of which include survey data, paradata, and behavior codes. We investigate how the question-answer interaction unfolds across conversational turns, accounting for the nesting of answers within interviewers, respondents and questions. Preliminary analyses show that the probability of obtaining an adequate response increases by approximately 10 percentage points (OR=1.5, $p<0.001$) in both studies if interviewers read the question exactly as worded, as opposed to reading the question with changes. Our analyses also suggest that there is significant variability in how this interaction unfolds at the question level (ICC=0.33). We will examine how this interaction differs across different types of questions. The paper concludes with implications of our findings for standardized interviewing, rapport building, and interviewer training.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Methodological Brief: Remind Me Again? Prompting and Reminding to Increase Response Rates

Date Me? An Experimental Examination of Including a Deadline on Survey Communications

Rebecca Powell *RTI International*

Emily Geisen *RTI International*

Marshica Stanley Kurtz *RTI International*

Murrey Olmsted *RTI International*

Survey researchers are continually trying to increase response rates to improve the robustness of survey estimates. One way they attempt this is by including information about the timeline of the survey in respondent communications. Two common approaches are (1) including a “deadline” when the survey will close, or (2) stating that respondents should return the survey “as soon as possible” (ASAP). Each method is intended to convey a sense of urgency to respondents and convince them to respond quickly. The literature indicates mixed results when comparing the use of ASAP versus a deadline. Some studies found an increase in response rates when a deadline was included compared to no deadline (e.g., Porter and Whitcomb, 2003; Martin, 2009), while others found no differences (e.g., Bouffard et al., 2004; Dillman, 1991). In Dillman and colleagues’ recent discussion on this topic (2014), they suggest using “as soon as possible” for most communications and reserving a deadline or the term “approaching soon” for the final mailing; however, they do not provide empirical evidence of the success of this approach. In this research, we test their recommendation with the 2017 Best Hospitals Physician Survey. We randomly assigned 4,700 physician specialists to one of three groups: (1) deadline mentioned in all four letters, (2) ASAP mentioned in all four letters, and (3) ASAP mentioned in the first three letters with the deadline mentioned in the final mailing. We compare response rates of the three groups at the end of data collection, as well as how quickly the surveys are returned. Results of this research can be used to inform survey researchers about best practices for conveying timeline information to respondents in order to increase response rates.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Methodological Brief: Remind Me Again? Prompting and Reminding to Increase Response Rates

Testing the Impact of the Type of Mail Used on Augmenting Response Rates for a Leave-Behind Questionnaire in a Face-to-Face Survey

Daniel Lawrence *NORC at the University of Chicago*
Erin Burgess *NORC at the University of Chicago*
Ned English *NORC at the University of Chicago*
Katie Archambeau *NORC at the University of Chicago*
Colm O'Muircheartaigh *Univ of Chicago NORC*

With declining response rates becoming the new norm, it is increasingly imperative to identify survey methods and protocols that increase respondent cooperation. For the recently concluded third wave of the National Social Life, Health, and Aging Project (NSHAP), NORC conducted an experiment to test whether (a) a mail follow-up to telephone prompting boosted response rates for a paper questionnaire, and (b) whether the type of mail used for a sequence of reminders had a differential impact. The NSHAP interview consisted of a 90-minute, in-person questionnaire and physical measures administration. At the conclusion of the interview, respondents were given a supplemental leave-behind questionnaire (LBQ) to return in a prepaid envelope. If respondents did not return the LBQ within three weeks, they were contacted by telephone in a sequence of up to three separate prompts to return it. (a) The mail follow-up produced consistent increases in response rates: about 19% of nonrespondents compared to 2% for the control group. (b) In the initial mail treatment, one group received their materials via regular mail and the other group received their materials via FedEx (the control group received no materials). For this initial treatment, receiving materials via FedEx produced significantly higher returns than via regular mail: 23% vs 14%. Different sequences of follow-up were tested subsequently, but while response rates between the experimental groups converged over time, the group receiving the initial FedEx treatment continued to outperform the regular mail treatment group. The paper will describe the full complement of treatments, present time-sequence results for all groups, and conclude by considering the application of these results to other studies and populations. Our research is relevant to practitioners of face-to-face and multi-mode surveys.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Methodological Brief: Remind Me Again? Prompting and Reminding to Increase Response Rates

Investigating the Effects of Survey Links on Response Rates

Raeal Moore *ACT, Inc*

Emily Uhl *ACT, Inc*

Research on how survey links affect response rates is limited. Placing the URL at the bottom of the email message improved response rates (Kaplowitz et al., 2012) even though placement at the top reduces the need to scroll to view the link (Couper, 2008; Dillman, 2014). No research on link type or quantity has been documented. We conducted an experiment to evaluate the effects on survey response rate and data quality of an emailed survey invitation's URL location, type, and quantity. A random sample of 29,280 September 2016 ACT test-takers were randomly assigned to experimental groups using a 2 (survey link location) x 2 (survey link type) research design. The survey link was placed either at the top or bottom of the invitation message. The type of link was either the hyperlinked word "survey" or the survey software platform's customary survey URL. An additional 14,640 students were randomly selected and assigned to two groups to determine what impact two types of links in combination had on response rates. One group received the word "survey" hyperlinked to the online survey on the top of the invitation and the survey software platform's customary survey URL on the bottom. The second group received links in the reversed location. Total response rate, across the two studies, was 10%. Providing the survey software platform's customary survey URL at the top yielded the highest response rate. When both types of links were provided, placing the customary survey URL at the top and hyperlink at the bottom yielded the highest response rate. Providing both types of survey links increased response rates relative to providing only one link. Other results showed no differential impacts on key outcomes. These results provide valuable insights to the design and quantity of survey links in invitations.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Methodological Brief: Remind Me Again? Prompting and Reminding to Increase Response Rates

Assessing the impact of web option for mothers of new children using the Tailored Design Method.

Kurt Johnson *RTI International*

Thomas Duffy *RTI International*

The Tailored Design Method (Dillman, Smyth, Christian, 2014) is a well-established methodology for maximizing response for self-administered mail surveys. Current prevailing adaptations call for a mode first design where the primary mode (mail) is used, with the addition of the option mode (web) as a final contact, and as is often the case a non-response follow-up call attempt. As survey demographics change to those who are more accustomed to digital communications, it is important to continue to test the value of this approach to data collection. To assess, we have developed an experimental design that randomizes the invitation to participate in a web survey across the four primary points of contact (prenotification, first survey mailing, second survey mailing, and final survey mailing). The population of this study is a cohort of mothers of children born within 2-months of the invitation to participate, and closely replicates the CDC PRAMS methodology. This primary goal of the experiment is to introduce the web option to potentially capture more women via self-administered mode before going to CATI. Results will indicate the impact of providing a mode-option earlier in the protocol on both cooperation and response, as well as potential selection bias. That the study uses a population of mothers of newborns, the results will also provide evidence of the effectiveness of mode-option across more hard to reach populations.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Methodological Brief: Remind Me Again? Prompting and Reminding to Increase Response Rates

Potential Unintended Consequences of an Email Reminder Strategy for a Household Survey with an Address-Based Sample Frame

Sarah Grady *National Center for Education Statistics*
Cameron McPhee *American Institutes for Research*

The potential to increase survey response by sending email reminders to sampled cases is seductive because email messages are inexpensive to send. However, privacy concerns may hinder respondents from providing email addresses within survey instruments or worse-may scare them away from responding to the survey entirely. This paper presents results of a 2016 email experiment ($n = 35,000$) which was designed to understand the impact of asking for email addresses in a two-stage household survey so that an email reminder strategy could be evaluated for future collections. Half of the respondents to the screener stage of the web survey were asked to provide the email address for the adult in the household who would be asked to complete the second-stage survey. Most of the time, the screener respondent was also asked to be the respondent for the second stage of the survey, but less than a third of the time, the screener respondent was asked for the email address of another adult in the household. Though email addresses were not used for nonresponse follow-up, the experiment yielded data about item nonresponse to the screener email question and about screener breakoffs, which lead to unit-level nonresponse for the second stage of the survey. Additionally, respondents who provided email addresses were sent a “thank you” email, thereby providing data about the number of email bouncebacks, which indicates the quality of the email addresses provided by respondents. Results indicate that, for the most part, the risk of unintended consequences of an email strategy is relatively low, with a low breakoff rate and relatively high item response and data quality. However, the authors found that asking for email addresses of other household adults yielded lower item response to the email question and appeared to encourage proxy responses to the second stage of the survey.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Methodological Brief: Remind Me Again? Prompting and Reminding to Increase Response Rates

Examining Phone Follow-up Effort in School Recruitment

Yan Wang *American Institutes for Research*

Corey Sinser *AIR*

Many school-based survey recruitment efforts include a phone follow-up component. Given the limited resources, it is challenging to determine the timing of the phone follow-up and the number of attempts an interviewer should make. To help plan for future school-based recruitment, we examined the outcome of our phone follow-up efforts during the recruitment for the ED School Climate Surveys (EDSCLS) national study. EDSCLS is a suite of survey instruments developed for schools, districts, and states by the National Center for Education Statistics (NCES) of the U.S. Department of Education. We collected data from a representative sample of schools across the United States to provide benchmarking for EDSCLS users. The recruitment began in the spring of 2016, with an initial mail packet. The first round of telephone follow-up was carried out from mid-May to the end of June with a total of about 590 schools. The second round of phone follow-up started in early August and lasted to early October and targeted about 230 new schools in addition to the schools that had not responded in the first round. Initial findings showed that on average it took 60 days and 14 call attempts to receive a participation decision from a school. After two rounds of phone follow-up, 51% of the schools provided us with a decision, while only 17% were able to have a decision after the first round. Among the schools that provided us with a decision, 42% were participants after two rounds of phone follow-up, while 72% were participants at the end of the first round. We also examined the relationship between school characteristics and the degree of phone follow-up efforts needed. We found that it took more attempts and longer period to recruit rural or urban schools compared with town or suburban schools.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Methodological Brief: Remind Me Again? Prompting and Reminding to Increase Response Rates

More Harm than Good? An Experimental Approach to Examining the Value of Evening and Weekend Calls

Casey A Easterday *HealthPartners Institute*

Jennifer M Renner *HealthPartners Institute*

Steve E Asche *HealthPartners Institute*

Kayla Dean *HealthPartners Institute*

Jeanette Y Ziegenfuss *HealthPartners Institute*

In the era of declining response rates, survey centers increasingly seek solutions to complete more surveys and reduce costs. For example, some shops staff evening and weekend hours to reach more respondents outside of business hours. However, staffing and financing these less desirable shifts can burden survey centers, especially smaller shops, and the benefit of staffing off-shift hours is often based on conventional wisdom rather than empirical evidence. To evaluate the value of staffing evening and weekend hours, we performed a full-factorial experiment in which 2,430 residents of MN and Western WI were randomized to one of eighteen treatments and received their first call attempt within one of three times of day (AM, PM and evening), and on one of six days (Monday through Friday and Saturday) to complete a telephone survey on their family's behavior and nutrition. We found no evidence to suggest evening or weekend calls increased contact or completion rates, and found evening calls increased refusal rates. Moreover, off-shift hours did not increase variation within the resultant sample; respondents completing the survey at different times of day or on different days did not vary systematically with respect to self-reported demographics.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Partisanship and Ideology in the 2016 Election

The Impact of Partisanship on Polling: Partisan Non-Response Bias and the 2016 Presidential Election Polls

Joshua D. Clinton *Vanderbilt University*

John Lapinski *University of Pennsylvania*

Partisanship is a powerful predictor in politics – affecting voters’ participation and opinions. We show that it also has an underappreciated impact on attempts to survey the current state of an election contests because of variability in the propensity of partisans to participate in election polling surveys. Using over 500,000 respondents from the NBC News/Survey Monkey tracking survey – including more than 1,000 per day on average -- we construct daily estimates of national opinion and show that there are sizable and influential variations in the partisan composition of respondents even after weighting to create demographically balanced daily samples. Moreover, we show that this is not a phenomena that only affects opt-in internet surveys; the same fluctuations in partisan composition post-weighting routinely occur in the non-internet polls that were regularly conducted throughout the campaign. To the extent that partisans differentially participate in public opinion polls, our findings have important and provocative implications for how we report and interpret changes in opinion over time. This is especially true when the differential participation is a consequence of campaign events for which the polls are being used to assess.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Assessing the Health of America: Values, Beliefs, Knowledge and Behaviors
Prescription Painkiller Use in America: Public's Views of the Epidemic and Personal Use
Bianca DiJulio *Kaiser Family Foundation*
Bryan Wu *Kaiser Family Foundation*
Scott Clement *The Washington Post*
Emily Guskin *The Washington Post*

Every day, 78 Americans die from an opioid overdose (including heroin and prescription pain relievers), reflecting a quadrupling of opioid-related deaths since 1999. In order to better understand the public's perception and proximity to this national epidemic, this paper will use data from nationally representative surveys conducted in 2015 and 2016 to examine what the public knows and understands about the epidemic, how much they are following the issue, and their views on how to address it. It will also explore personal experience with addiction and misuse and how those experiences shape overall views of the issue. In addition to surveys of the general public, the paper will include forthcoming representative survey data that takes an exhaustive look at how individuals personally using strong prescription painkillers feel about the epidemic and efforts to stop it, as well as their reasons for use, concerns about addiction, and personal circumstances.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Assessing the Health of America: Values, Beliefs, Knowledge and Behaviors
Public Opinion on the Zika Virus Outbreak: Americans' Knowledge and Worries

Bryan Wu *Kaiser Family Foundation*
Bianca DiJulio *Kaiser Family Foundation*

Our organization has been tracking public opinion on the Zika virus outbreak since February 2016, when the World Health Organization first declared the virus to be a public health emergency. This project focuses on the American public's knowledge of, and worries regarding, the Zika virus outbreak over the course of 2016, leveraging public opinion data on previous virus outbreaks like the Ebola outbreak of fall, 2014 to serve as comparison when available. Surveys were conducted using a nationally representative random digit dial telephone sample, with interviews conducted by landline and cellphone. As expected, we observe differences in both knowledge of key facts related to Zika and worries expressed about the virus by level of awareness of Zika-related news coverage – generally, those who have heard “a lot” about the virus are more aware of key facts and express higher degrees of worry than those who have heard “only a little.” We additionally explore interactions that affect knowledge of and worry for the Zika virus. Specifically, we examine the interplay between individuals' levels of awareness of Zika-related news coverage and risk factors for Zika – being female, pregnant or trying to get pregnant, and geography (both in terms of mosquito prevalence and regions of the U.S. most at-risk for mosquitoes carrying the virus). We expect to demonstrate that level of awareness of coverage is a more important factor associated with Zika knowledge than is being part of a group for which Zika might be a particularly salient topic. This project provides useful insight, as health officials warn that Zika will most likely become endemic to the United States and a threat for the foreseeable future.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Analyzing & Managing Cost With TSE

Assessing Impact of Data Collection Level of Effort and Design Features on Bias and Cost

Pamela Giambo *Westat*

W. Sherman Edwards *Westat*

Douglas Williams *Westat*

Lynn Langton *Bureau of Justice Statistics*

Our paper is based on data from a large-scale, experimental mail survey about victimization. The goal of the study is to develop a lower cost, companion survey for the National Crime Victimization Survey (NCVS), sponsored by the Bureau of Justice Statistics (BJS). More than 200,000 households were selected in 2015 and in 2016. The survey tested various approaches to improve response, including variation of survey content in year 1, and variation of incentives and delivery mode in year 2. Research questions include whether experimental treatments reduced bias and improved representation from diverse sub-populations sufficiently to justify the additional cost. A 2015 experiment varied the use of additional survey content to improve survey response by increasing content saliency. The additional content covered perceptions about safety and police performance. The hypothesis was that non-victims would have lower response propensity because the victimization survey was not relevant to them. By including content relevant to all households at the beginning of the survey questionnaire, the study designers hoped to improve response propensity among non-victims. In 2016, the study varied incentives (\$2, \$1, and no incentive) and mail delivery method (first-class mail vs. Federal Express). The question here is whether any reduction in bias is large enough to justify the additional expense. Given the scope of the survey, we have considerable power to assess these differences. A final experiment used a partial overlap of the sample between 2015 and 2016. The rationale for this was that the estimates of change would yield higher levels of precision for the overlap sample. One concern is that response propensity may change for households that already received a survey in 2015. If this propensity changes differentially for victims compared with non-victims then the overlap could introduce differential bias.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Assessing the Health of America: Values, Beliefs, Knowledge and Behaviors

A mixed methods approach to explain the declining female life expectancy in US counties

Charbel El Bcheraoui *University of Washington*

Kaylin Bolt *University of Washington*

Casey Johanns *University of Washington*

Erin Palmisano *University of Washington*

Aubrey Levine *University of Washington*

Laura Dwyer-Lindgren *University of Washington*

Joseph Dieleman *University of Washington*

Ann Madhavan *University of Washington*

Ali Mokdad *University of Washington*

The United States ranked 40 worldwide with 79.1 in life expectancy for 2015. Moreover, large disparities in life expectancy between US counties exist for both men and women. Since 1980, female's life expectancy declined in 13 counties and increased by less than two years in 251 counties. Risk factors such as obesity, high blood pressure, and smoking explain some, but not all of these trends. However, a more comprehensive look at the drivers requires a combination of quantitative and qualitative approach. Based on several risk and socioeconomic factors, we selected ten counties to investigate the factors driving health and mortality. A pair of counties were selected from five states. Each pair included one county with increasing, and one county with stagnant or decreasing life expectancy trend. We conducted face to face interviews using semi-structured topic guides to investigate factors possibly driving health and mortality within a socio-ecological framework. We targeted key informants from relevant sectors, and female residents. Recorded interviews were transcribed and analyzed through recursive abstraction. The qualitative methods identified many drivers of health that could explain declining life expectancy among US females in addition to those identified by the quantitative analyses. Major themes were reported by more than 80% of respondents. A predominant culture of health where county residents valued their health status and tended to live a healthier lifestyle, a strong and growing economy, and high quality accessible health care systems sprawled in counties with increasing life expectancy. On the other hand, loss of major industries and consequently high paying jobs, a culture of tolerance of health problems, limited resources, and poor and inaccessible health care systems impeded counties where life expectancy decreased on did not improve. Respondents explained how these factors trickled down to the inter-personal and individual level factors, in regards to health related behaviors.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Analyzing & Managing Cost With TSE

Rising costs in survey administration: Maintaining long-term trends without blowing the budget

Sarah L Hagge *Minnesota Department of Health - Health Economics Program*

Alisha BainesSimon *Minnesota Department of Health - Health Economics Program*

Kathleen Thiede *Univ. of Minnesota, SHADAC*

Giovann Alarcon *University of Minnesota, SHADAC*

Kendal Orgera *Minnesota Department of Health - Health Economics Program*

Karen A Turner *University of Minnesota, SHADAC*

Survey administration costs have climbed in response to decreasing response rates, restrictions in automatic dialing, and sampling strategies to reach shrinking populations (e.g., people without health insurance). Researchers face the issue of maintaining robust estimates and long-term trends with budgets that do not meet cost demands. This paper simulates the cost reduction strategy of administering two shorter survey forms and explores the extent to which population estimates change under this strategy. Every two years, the state of Minnesota conducts the Minnesota Health Access Survey, a population-based dual-frame telephone survey on health insurance coverage and access. Results of the survey are used to determine the size of the uninsured population and track changes in access to healthcare among all Minnesotans. The target sample size for the survey has historically been 11,000, though subgroups such as the uninsured make up less than 5% of that. Exploratory results indicate that differences in estimates may vary according to the sample. For example, respondents were asked whether they had forgone care due to cost. Overall, 85.5% (SE = 0.7%) of people with group health insurance reported they had not forgone care due to cost. When a two form design was simulated by randomly dividing people with group health insurance into two subgroups, estimates of 85.2% (SE = 1.0%) and 87.1% (SE = 0.9%) were obtained. Neither estimate was significantly different from the total sample, but this highlights the variability that could arise in estimates. Given initial evidence, bootstrap resampling methods will be applied to evaluate bias and variability of a range of estimates. The original estimates from the full 2015 survey will be considered the criterion against which the sample conditions will be evaluated. Bias, standard error and root mean square error between the criterion and study conditions will be calculated to summarize error.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Partisanship and Ideology in the 2016 Election

What Was Bad Is Now Good, What Was Good Is Now Bad: Pinpointing Changes in Partisans' Views of National Conditions Under a New President

Jeffrey Jones *Gallup*

Americans' assessments of the situation in the country – the health of the economy, satisfaction with the way things are going, even their own personal finances – are colored greatly by the match between the party of the president and their own partisanship. When the party of the president changes, Republicans' and Democrats' views of national conditions often flip with the side formerly more negative becoming more positive and the side formerly more positive becoming less so. Gallup polling, in particular its daily tracking survey, can help pinpoint exactly when these partisan changes in national evaluations occur under Donald Trump – when he won election, at the time he took office, or after he has been in office for a while, and whether the timing varies by item. Gallup has already observed movement on some measures after Trump was elected, but others may not change until he takes office. Gallup can compare the data for Trump with similar daily tracking data collected when President Obama replaced President Bush, as well as for prior presidential transitions when possible.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Analyzing & Managing Cost With TSE

Total Error in a Big Data World with Applications to the Residential Energy Consumption Survey

Ashley Amaya *RTI International*

Paul Biemer *RTI International*

David Kinyon *Energy Information Agency*

The use of big data to supplement or replace survey data is growing. In the world of escalating survey costs, falling coverage and response rates and the rising potential for nonsampling bias, big data offers an arguably less expensive, less burdensome, and timelier alternative to produce a variety of statistics from flu prevalence to consumer expenditures to neighborhood characteristics. However, big data is not without its weaknesses; it suffers from its own sources of error, access challenges, and confidentiality concerns. The AAPOR Task Force on Big Data called for researchers to evaluate the quality of big data using an approach similar to the total survey error framework (Japec et al 2015). By understanding the primary sources of error, we may create data processing and analytic methods to minimize or correct it. When supplementing survey data with big data, understanding the primary error sources of each data set allows us to capitalize on the strengths from each to create a stronger, combined product (Biemer 2016). In this paper, we demonstrate an approach to apply the total error framework to evaluate big data quality using dwelling unit area (square feet) from three databases. We then identify ways that these data may be used to supplement survey data from the Residential Energy Consumption Survey (RECS 2015) and investigate whether combined estimates produce lower error rates than are possible from either individual source.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Assessing the Health of America: Values, Beliefs, Knowledge and Behaviors

RWJF's American Health Values Survey: A New Typology

Larry Bye *NORC at the University of Chicago*

Carolyn Miller *Robert Wood Johnson Foundation*

Alyssa Ghirardelli *NORC at the University of Chicago*

Angela Fontes *NORC at the University of Chicago*

RWJF's Culture of Health (CoH) vision embraces a broadly integrated and comprehensive approach to health—one where well-being lies at the center of every aspect of American life. To better understand how Americans differ with respect to the CoH vision, the Foundation commissioned The American Health Values Survey (AHVS), a national survey of US adults. A probability-based sample of over 10,000 adults was analyzed to develop a typology that segments U.S. adults based on their health values and beliefs (CASRO Response Rate 22.4%). Survey measures at an individual level captured the importance of health as a life concern; how health is defined; and self-efficacy about prevention, care-seeking and condition management. At the societal level, data was collected on beliefs about equality of opportunity, health equity, the social determinants of health, health disparities, the role of government and civic engagement on health. Six population groupings, or segments, were identified. Three segments included people who support an active government role in health, named Equity Advocates, Health Egalitarians and Committed Advocates. There is clearly the potential to mobilize all three of these segments in efforts to improve population health and health equity in the US. Two segments were characterized by people who were skeptical toward government involvement in health, Self-Reliant Individualists and Disinterested Skeptics, and one segment was characterized by people with conflicted attitudes toward government in health, Private-Sector Champions. For those in more skeptical segments, a focus on building healthier communities at the local level may garner support as long as there is private-sector leadership for the effort. We will present how American adults are highly divided in their health values and beliefs. These differences and divisions will also be explored through qualitative research in early 2017 to provide more insights into the implications for development of policy, program and communication strategies.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Assessing the Health of America: Values, Beliefs, Knowledge and Behaviors
Can Your Community Make You Healthy? Active Living Environments and their Impact on Residents' Well-Being

Diana Liu *Gallup*
Nader Nekvasil *901 F Street NW*

This study explores how community infrastructure impacts residents' healthy behaviors and their overall well-being. According to research from Gallup-Healthways, communities with active living environments—those communities that invest in bike paths, parks, walk-ability and public transit—have residents with better outcomes in key elements of well-being. For this study, Gallup-Healthways merged publically available open data from Walk Score® and ParkScore® with data from the Gallup Daily Poll—a nationally representative poll of U.S. adults on various political, economic and well-being topics. In addition to the bike-ability, walk-ability and quality of transit infrastructure within communities, this study will also examine how access to organic grocery stores, proximity to fast food restaurants, real estate stability and other economic factors can greatly impact subjective well-being and daily emotions among residents.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Partisanship and Ideology in the 2016 Election

Partisan Stability and the 2016 presidential campaign

Bradley Jones *Pew Research Center - Washington, DC*

Alec Tyson *Pew Research Ctr.*

Jocelyn Kiley *Pew Research Ctr.*

John Oliphant *Pew Research Center*

Partisanship is one of the most fundamental attitudinal constructs shaping individual level attitudes and behavior. Political scientists have long documented the stability of individual-level partisan identification (Green, Palmquist, and Schickler 2004), but to observe that it is stable is not to claim that it never changes. Indeed, partisanship moves in the aggregate in a way that cannot be attributable solely to generational replacement (MacKuen, Erikson, and Stimson 1989), and it has been shown to move even in short time frames (Smidt 2014). This paper analyzes a high-quality nationally-representative panel dataset that traces individual-level changes in party loyalty before, during and after the 2016 presidential campaign. We show how assessments of the candidates, disagreement on particular issues, and demographics shape partisan identity. In addition to standard partisanship items, the panelists responded to several items designed to assess their attachment (or lack of attachment) to the major parties in the United States. This paper contributes to our understanding of an important moment in American politics, but it also sheds light on some of the underlying factors that shape partisanship more generally.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Analyzing & Managing Cost With TSE

Assessing Error in Health Insurance Estimates: A Total Survey Error Perspective and Demonstration

Matt Jans *UCLA Center for Health Policy*
Todd Hughes *UCLA Center for Health Policy Research*
Tara Becker *UCLA Center for Health Policy Research*
Kevin McLaughlin *UCLA Center for Health Policy Research*
Royce Park *UCLA Center for Health Policy Research*
Emilia Peytcheva *RTI International*
Andy Peytchev *University of Michigan*
Jill Dever *RTI International*
Jamie Ridenhour *RTI International*
Ismael Cervantes *Westat*
J. Michael Brick *Westat*
Hongjian Yu *UCLA Center for Health Policy Research*
Yueyan Wang *UCLA Center for Health Policy Research*
Gerald Kominski *UCLA Center for Health Policy Research*
Ninez A Ponce *UCLA Center for Health Policy Research*

Error exists in all estimates from survey data but it is often very difficult to study without specialized evaluations and clear counterfactuals (whether by experimental design or sophisticated statistical analyses). For example, data on respondents and nonrespondents are needed to disentangle nonresponse error from other survey errors. Such complexities make it impossible to assess even one aspect of Total Survey Error (TSE) in many limited-resource contexts. This paper demonstrates several methods of assessing aspects of TSE in estimates of the uninsured population and related health insurance status estimates from the California Health Interview Survey (CHIS). Although past research on CHIS shows no need for concern about nonresponse bias in these estimates, increasing nonresponse rates introduce additional error risk. First, we benchmark CHIS estimates of the uninsured population against estimates from California samples of the American Community Survey (ACS), National Health Interview Survey (NHIS), Current Population Survey (CPS), and Behavioral Risk Factor Surveillance System (BRFSS). We accomplish this with a TSE-based checklist tool that takes measurement and representation errors into account simultaneously. Second, the magnitude of nonresponse adjustments and other characteristics of CHIS weights are used to assess the presence of nonresponse bias in estimates before adjustment. Initial findings show that CHIS obtains uninsured rates similar to BRFSS, but different from other federal surveys, suggesting an effect of sample design or data collection mode. The presence of large nonresponse adjustment factors in CHIS suggests that nonresponse is accounted for in weighting. Through examples, we discuss a continuum of error assessment methods that can be used in a range of survey environments.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Partisanship and Ideology in the 2016 Election

How Independent are Registered Non-Partisans? Exploring Party Leanings among California's Growing Number of Independent Voters

David R Kordus *Public Policy Institute of California*

In the 2016 presidential campaign, a longtime independent seriously contended for the Democratic Party's nomination, and the winner of the Republican Party's nomination and the presidency had recently been registered as an independent. In California today, nearly one in four voters are registered as independents (officially "no party preference") – a share of voters that has more than doubled in the past twenty years. As independent voter registration increases, pollsters and public opinion analysts may leave much unexplained by treating this large group of voters as homogeneous. Previous research into party identification has shown that most survey respondents who identify as independents will say they lean toward one of the major parties when asked. Furthermore, the preferences and voting behavior of party leaners look a lot like the preferences and voting behavior of partisans. Should we expect similar patterns among voters who officially register as independents, or are they a distinctly more "independent" group? Using survey data from the Public Policy Institute of California's periodic statewide surveys, this paper examines survey responses from registered independents during 2016 and explores how opinions among independents compare to those of registered party members. Candidate and policy preferences, attitudes toward the major parties, and political ideology are considered, with an objective of gaining a more thorough understanding of the large and growing number of registered independents. This examination concludes that California's registered independents are not a homogeneous group and that independents predictably share views with the political party they say they lean toward. This finding demonstrates the value of considering the partisan leanings of registered independents in survey analysis, which can help to illuminate survey findings and trends of increasing non-partisan registration in California and elsewhere.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Analyzing & Managing Cost With TSE

Deploying a Total Survey Error (TSE) and Total Survey Quality (TSQ) Assessment of the AmeriSpeak® Panel

Vicki Pineau *NORC at the Univ of Chicago*

Paul JLavarakas *Independent Consultant*

J. Dennis *NORC*

Many survey researchers are familiar with the Total Survey Error (TSE) framework. Not so many are familiar with the Total Survey Quality (TSQ) approach (cf. Biemer, 2011; Lyberg, 2015). These scholars have suggested that TSE is of greatest value to survey researchers when it is used “as part of a larger design strategy that seeks to optimize TSQ.” TSQ encompasses survey design, survey implementation, survey evaluation, and the effects of survey errors (both bias and variance) on survey analyses. In 2016, a systematic and comprehensive review of the NORC’s probability-based AmeriSpeak® Panel is being carried out that is gathering information on what is known about AmeriSpeak’s Errors of Representation (e.g., errors related to coverage, sampling, nonresponse, and weighting) and Errors of Measurement (e.g., errors related to questionnaire design, interviewers, respondents, and the use of multiple modes for data collection) – culminating in a quality profile. It also is gathering information about a host of operational factors that are related to and affect the functioning of AmeriSpeak®. The presentation will address the application of the TSE and TQS approaches in evaluating the progress of NORC’s AmeriSpeak® Panel during the first two years of the panel’s existence.

May 19th, 2017
4:15 PM - 5:45 PM
Concurrent Session E

Partisanship and Ideology in the 2016 Election

Ideology Versus Party Identification: Which Measure is More Stable?

Sarah Cho *SurveyMonkey*

Jon Cohen *SurveyMonkey*

When it comes to the stability of political ideology and partisan affiliation, results have been mixed. Some researchers have found that ideology is more stable than party identification, while others have found that partisan affiliation is more stable than ideology - even using the same dataset. In this study, we examine which is more stable: political ideology or party identification. After individuals completed a user-created online survey on the SurveyMonkey platform, we asked respondents to take one more survey related to politics. Upon completing this political poll, respondents were then asked to join a non-probability panel and recontacted again at a later date. In this presentation, we will examine whether party identification or political ideology is more stable over time within these individuals that completed the political survey and joined the panel. We will also present on the stability of other views, including candidate support and presidential approval.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Geographic Information Systems (GIS) Methods and Technology for Survey Research (Panel)

Geographic Information Systems (GIS) Methods and Technology for Survey Research

Stephanie Eckman *RTI International*

Andrew Anesetti-Rothermel

Stephanie Eckman *RTI International*

Brad Edwards *Westat*

Marsha Hasson *Westat*

Andrew Anesetti-Rothermel *The Schroeder Institute for Tobacco Research at Truth Initiative*

Victoria Vignare *Westat*

Susan Genoversa *Westat*

Kyle Fennell *NORC at the University of Chicago*

Ned English *NORC at the University of Chicago*

Michael Keating *RTI International*

Julia Brinton *RTI International*

Robert Furberg *RTI International*

Jennifer Cantrell *Evaluation Science and Research at Truth Initiative*

Peter Herman *NORC at the University of Chicago*

Morgane Bennett *Evaluation Science and Research at Truth Initiative*

Elizabeth Hair *Evaluation Science and Research at Truth Initiative*

Donna Vallone *Evaluation Science and Research at Truth Initiative*

With the continued growth in the Geographic Information Systems (GIS) industry, GIS is playing a growing and invaluable role in many aspects of survey research. GIS has been used traditionally for sample plan development and mapping of data however GIS offers tremendous potential for additional advanced tasks. The use of expanding GIS industry technologies and geographic based statistic methods offer new ways to analyze and understand data trends for survey research projects to reveal previously unrecognized spatial patterns. The use of locational based data collection technologies, such as GPS, allow for survey data to be collected and placed in a geographic context. This panel will present a collection of papers on advanced applications of geospatial methods with a focus on location based tools for survey research.

Demographic Disparities in the Tobacco Retail Environment in Washington, DC: A Districtwide Spatial Analysis

Andrew Anesetti-Rothermel

.Research has shown that neighborhoods with higher racial/ethnic concentrations and lower SES tend to have a higher density of tobacco retailers. This replication study utilizes a spatial analytical framework to evaluate the racial/ethnic and socioeconomic disparities in neighborhoods (i.e., census tracts) in Washington, DC. Unlike the original study, an adaptive-bandwidth kernel density estimation approach was used to calculate tobacco retailer density (i.e., number of outlets per 1,000 people) within in each census tract. A districtwide density surface was constructed from tobacco retailers (n=1,955) identified by unique retailer codes within a national business directory. Zonal statistics were used to derive mean

density estimates for each census tract. Sociodemographic data for minority neighborhood racial /ethnic composition and socioeconomic disadvantage conditions were obtained from the 2010-2014 American Community Survey. Descriptive statistics and geovisualization assessed the distribution and spatial patterning of tobacco retailer density across the district. Additionally, spatial autocorrelation within the spatial pattern was assessed via a Global Moran's I test statistic. Spearman's correlations between the demographic characteristics and tobacco retailer density were computed with and without accounting for spatial autocorrelation. Log-linear bivariate and multivariate ordinary least square (OLS) regression and spatial regression models were then fitted to predict tobacco retailer density. Multivariate models controlled for all demographic variables as well as population density. OLS and spatial models were compared using Akaike Information Criterion. Census tracts had a mean density of 4.68 tobacco retailers per 1,000 people (SD=3.24). Geovisualization revealed significant global spatial autocorrelation in the spatial pattern of tobacco retailer density (Global Moran's I=0.71, p=0.001). Similar to the original study and contrary to existing research, no demographic characteristics were associated with tobacco retailer density, which suggests that there are no demographic disparities in the tobacco retailer environment in Washington, DC.

Using GPS to Detect Falsifiers: Some Nuts and Bolts

Marsha Hasson

Falsification has been a longstanding concern in surveys, and the face-to-face mode is especially at risk because the interviewer's work is performed with only light remote supervision. GPS has emerged as a new tool to detect falsifiers in the field, by matching the interviewer's location with the location of the respondent's home at the time the interview took place. This new tool has one major advantage over other forms of detection: it can be applied nearly universally across all interviews, rather than to just a sample. It scales up. This paper will describe how GPS data are used to support identification of potential falsification at Westat. A mobile or laptop device logs the location of the interviewer throughout the day, including travel and stationary activities. Depending on approach, this information is sent to the home office in near real-time or transmitted by the interviewer following interviews. Immediately upon receipt, the GPS information is combined with case data and various algorithms are run to determine if the interviewer was at the home of the respondent at the time the interview conducted. The supervisor is immediately alerted of any issues and provided with the case data, including any EROCs that might explain a change of location at the respondent's request. The supervisor begins an investigation that may include a review of the interviewer's route that day, caseload, records of contact attempts, CARI data, and work history. We provide several examples of these investigations that separate false positives from confirmed falsifiers. We will also provide some level of effort information and compare the costs and benefits of using GPS versus CARI and re-interview methods as a tool for detecting falsifiers.

Using GPS traces to Evaluate Interviewer Efficiency

Kyle Fennell

A key feature of NORC's AmeriSpeak panel is the use of field interviewers to enroll panelists as part of a non-response follow-up (NRFU) for a sub-sample of addresses still pending after mail, web, and phone outreach. Because in person enrollment is expensive relative to other modes, it is important to enable interviewers to be efficient when working their cases. This paper introduces the structure of the NRFU effort and features of the case management system used by interviewers during the 2015 and 2016 AmeriSpeak sample building efforts. It then discusses how GPS data were logged as interviewers worked their cases and analyzes differences in how interviewers who were more efficient and effective

structured their work in the field when compared to their less effective peers. Finally, recommendations for systems, staffing, training, and management changes for future efforts are discussed.

Examining the Fidelity of Location-Based Dataset Linkage Using Fitbit Devices

Michael Keating

As technology increasingly becomes an integral part of everyday life, many individuals are choosing to use wearable technology such as Fitbit to track their daily physical activity and other health-related goals. Researchers would benefit from being able to remotely access the physical activity and health data, without meeting face-to-face with participants and avoiding the high cost of providing consumer wearables to participants for the study duration. The present study will seek to test the fidelity of remotely collecting physical activity, health-related data, and GPS location from participants without ever meeting them face-to-face. 30 participants will be recruited on Mechanical Turk Prime and asked to complete a short online questionnaire. They will be asked to connect their personal Fitbit device to an online third party software system. They will also be asked to allow researchers to access the IP address on their device in order to obtain location data. The present study will test a protocol whereby respondents connect their own personal consumer wearable device to an online data downloading software system, without the need to meet face-to-face with the research team, or provide them with study-owned devices. This protocol would allow for access to self-tracked physical activity data, location of respondents, and self-administered responses about health, all without the need to connect with participants in person.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Research on Collecting Information About Race and Ethnicity in the 2015 National Content Test (Panel)

Research on Collecting Information about Race and Ethnicity for the 2020 Census

Michael Bentley *U.S. Census Bureau*

Sarah Konya

Sarah Konya *U.S. Census Bureau*

Kelly Mathews *U.S. Census Bureau*

Rachel Marks *U.S. Census Bureau*

Julia Coombs *U.S. Census Bureau*

Nicholas Jones *U.S. Census Bureau*

In the fall of 2015, the U.S. Census Bureau conducted a census test of contact methods and questionnaire content that reached a nationally representative sample of 1.2 million households. One major focus of the 2015 National Content Test was to explore new methods of collecting information about race and ethnicity. This research aimed to refine our efforts to address known race and Hispanic origin reporting issues while improving data in three crucial areas: 1) increasing accuracy of reporting in the major Office of Management and Budget racial and ethnic categories; 2) collecting detailed data for myriad groups; and 3) obtaining lower item nonresponse rates. Households received one of 36 unique questionnaires where the format of the race and ethnicity questions, the available response options, and the wording of the question and instructions differed. Responses were then compared with responses to a telephone reinterview for a subset of cases to measure consistency. In this session, a panel of research experts from the Census Bureau will provide an overview of the methodology for the 2015 NCT research on race and ethnicity and will discuss results for each of the research dimensions.

Background and Methodology on the Census Bureau's 2015 National Content Test

Sarah Konya

The first paper about the 2015 National Content Test's race and ethnicity research will provide an introduction to the key research dimensions of the test and why we are examining these important topics. This presentation will highlight how the 2015 NCT research presents the critical opportunity to compare the success of different race and ethnicity question designs to determine how they perform in traditional paper designs as well as web-based data collection methods using the Internet and telephone response options. The complex sample of 1.2 million housing units across the United States was designed to ensure that the estimates from the test accurately reflected the nation as a whole and across a variety of demographic characteristics. This presentation will describe how the sample was constructed and how we effectively oversampled for key race and ethnic groups. Finally, the reinterview, which aimed to assess the accuracy of the various race and ethnicity question designs, will be discussed. The reinterview was designed to measure respondents' self-identified "true" racial and/or ethnic identity through a series of detailed questions and probes. The presentation will discuss how the reinterview was constructed and conducted as well as how the results were evaluated.

Evaluating a Combined Race and Ethnicity Question Format Compared with a Separate Question Design

Kelly Mathews

The second paper about the 2015 National Content Test's race and ethnicity research will discuss the methods and results for testing alternative versions of the race and ethnicity questions to improve question design and data quality, particularly through the use of separate questions about Hispanic origin and race or through a single question that combines all response categories. We will begin with an in-depth description of the different question format designs tested in paper and web-based modes and explore how the designs are related to the editing of response data. The presentation will illustrate how different question format designs performed through the examination of distributions, consistency scores, the amount of detailed information collected, and the frequency of multiple responses collected. We will discuss what the results indicate about utilizing either a Two Separate Questions approach or a Combined Question approach to collect the most accurate data on race and ethnicity.

Testing a Middle Eastern or North African Race and Ethnicity Category

Rachel Marks

The third paper about the 2015 National Content Test's race and ethnicity research will discuss the methods and results for testing the inclusion of a distinct category for respondents of Middle Eastern and North African (MENA) heritage in the United States. We will review the use of the new response option in the experimental design and discuss how editing of responses was affected by the inclusion of a distinct category for respondents of MENA heritage. Since such a group is new to the Census Bureau's race question, this presentation will describe how a working classification was developed for the development of example groups for this MENA category and for data editing. The presentation will illustrate how the inclusion or exclusion of a MENA category impacted reporting and how effective the category was in eliciting responses across various groups such as Lebanese, Moroccan, Syrian, etc.

Evaluating Alternative Instructions and Terminology for the Race and Ethnicity Questions

Julia Coombs

The fourth paper about the 2015 National Content Test's race and ethnicity research will discuss the methods and results for testing alternative versions of the race and ethnicity questions with different instructions and terminology to improve question design, data quality, and respondent understanding of the race and ethnicity question(s). Beginning in Census 2000, respondents have been permitted to report more than one race, and the Census Bureau continues to test ways to communicate this option to respondents. In this test, the usual instruction, "Mark [X] one or more boxes" was compared with a new instruction to, "Mark all boxes that apply..." and to "...note, more than one group may be selected." At the same time, the test continued to refine the terminology used to guide the respondent to answer the question; the use of "race or origin" was compared with the use of "race or ethnicity" or the general word "categories." This paper will illustrate how the instruction wording and question terminology affected the reporting of multiple racial and ethnic groups and will show how the use of data collected during the reinterview confirmed the consistency of responses.

Summary of Census Race and Ethnicity Research Findings and Next Steps

Nicholas Jones

The fifth paper about the 2015 National Content Test's race and ethnicity research will explore how the innovative methods and design of this test enabled the Census Bureau to implement and evaluate cutting-edge designs for eliciting and collecting detailed information for our myriad racial and ethnic

groups in the United States. This presentation will bring together the overall findings to discuss how, through all of these analyses, the Census Bureau has advanced the body of knowledge on effective ways to measure the complex racial and ethnic identities of our diverse Nation. The presentation will illustrate how each of the conclusions for the race and ethnicity research dimensions (separate or combined format; MENA or no MENA category; instructions and terminology) come together to inform an optimal question design to provide the most accurate and relevant race and ethnicity data possible about our changing and diversifying nation.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Methodological Brief: Interviewers, Quality Control & Sample Design
Using Census and Surname Data to Oversample Racial/Ethnic Minorities in DC: Lessons Learned
Jordon Peugh SSRS
Michael Bader *American University*

This paper examines the use of Census data and ethnic surname data to attempt to oversample racial and ethnic minorities in Washington DC. American University and SSRS conducted a pilot survey in DC to gather data about neighborhood attachment, respondent health, perceived safety, and trust in institutions within two distinct kinds of neighborhoods. The survey was conducted by mail using address based sample. Census data were used to identify neighborhoods that were “Quadrivial Neighborhoods” (tracts that contain at least 10% each of Whites, Blacks, Latinos, and Asians and not more than 50% of any group) and “Disproportionally Latino Neighborhoods” (tracts that contain at least 25% Latinos). To achieve better representation of ethnic minorities, the sample was further stratified within these neighborhoods to oversample households with Asian or Hispanic Surnames and that are in high density African American neighborhoods. For this paper, we compare the expected distribution of race and ethnicity within each sample strata to self-reported survey data. We found the actual returned sample included larger numbers of Asians and Whites and fewer Hispanics and African Americans than expected in almost every sample stratum. The sample design was impactful in increasing the numbers of Asian and African American respondents and in decreasing the number of White respondents. However, it yielded fewer Hispanic respondents than planned and lower representation of Hispanics in these neighborhoods than in Census data. This paper will present the expected vs. actual distribution of respondents by race/ethnicity within each sample strata; identify patterns of variation in expected vs. actual by neighborhood and by racial/ethnic group; and discuss implications for the use of Census and surname data in ABS sample design. As this was a pilot study, we will summarize what we have learned and how these data will help us improve the sample plan for future research.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Methodological Brief: Interviewers, Quality Control & Sample Design

Context and Interviewer Factors: The Relationships and Effects on Survey Data

Keisha Miles *Temple University Institute for Survey Research*

Nina Hoe *Institute for Survey Research - Temple University*

Background: It is widely known that interviewer characteristics have an impact on survey data in household survey projects. Studies have shown that interviewers can influence respondents' answers through their behaviors and physical characteristics. In addition, cluster effects at the neighborhood or block level are often present. While much is known about interviewer and cluster effects, the ways in which interviewer characteristics interact with neighborhood characteristics is much less known. This study examines how the interview environment and the respondents' experience interact with interviewer characteristics and experiences, as well as the survey response. Methods: For a household survey of 833 African-American males in a major U.S. city, 10 field interviewers from different backgrounds and locations approached a random sample of addresses to conduct interviews. Following each completed interview, interviewers answered 6 questions about the context, including evidence of "neighborhood disorder," the interview location, the presence of other people, the respondents' perceived level of resistance during the consent process and comfort during the interview, and finally, the interviewers' own comfort level during the entire process. Survey data was merged with the context and interviewer data. Results: Preliminary results reveal many important relationships. Respondents' comfort level during the interview was not related to measures of neighborhood disorder, location of the interview, or presence of others. However, interviewers' level of comfort was related to all of those same factors. In addition, local interviewers felt more comfortable during the interviews as compared to non-locals; however this had no effect on respondents' resistance during the consent process or comfort during the interview. Interviewer and contextual factors were related to patterns in survey data. Conclusions: This study reveals the salient ways in which environmental factors, and respondent and interviewer characteristics and experiences interact. Findings underscore the importance of attending to these interacting contextual factors and effects.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Methodological Brief: Interviewers, Quality Control & Sample Design
Dealing with Young Adults Who Won't Answer The Phone: A Strategy for Improving Sample Balance
Prior to Data Collection

Jason A Husser *Elon University*
Kaye Usry *Elon University*

Differential nonresponse, and the error it introduces, has long plagued survey researchers. In recent years, the problem has only intensified, particularly when it comes to reaching young adults by phone. Perhaps more than any other group, they are far less likely to answer telephone surveys. The most common strategy for correcting this problem has been to weight after data collection. However, particularly when it comes to election surveys, intensive weighting is not without drawbacks. Using a 2016 election survey of likely voters in North Carolina, we implemented a method to correct for differential nonresponse prior to data collection. Based on a previous sample of likely voters, we estimated the probability of survey completion according to age. Using these probabilities, we then estimated the probability of completion for every potential respondent in our list of registered voters, and drew a weighted sample based on those probabilities. This technique is unique from other forms of oversampling in that we used the probability of completing the survey by age, to estimate how many young voters we needed to attempt to contact. We find that when compared to a sample conducted without this adjustment, this technique results in an unweighted sample that more closely reflects the population, and in post-stratification weights with much lower standard deviations. We argue this technique is a promising area for future research, and a useful tool for other researchers attempting to survey young adults. We discuss the limitations of this correction, and provide detailed information about how to implement this practice, even without information about the probability of survey completion.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Methodological Brief: Interviewers, Quality Control & Sample Design
Method of CATI/CAPI Quality Control Using Paradata
Semen Kostin *Public Opinion Foundation*

Organizations conducting survey research aim to obtain data of most possible quality. In audio computer-assisted interviews, monitoring of interview audio records seems to be a widespread technique. However, a standardized control procedure of interviewers does not exist but its different aspects are being discussed (see Baker et al. (2013) for interviewer rating criteria, Tarnai (2007) for procedure description etc.). Due to limited resources, researcher usually has no possibility to monitor all audio recordings; at the same time there is a lack of attention to the selection of interviews to be monitored. This paper suggests a method of selection of interviews to be monitored in comparison to commonly used random selection of designated share of complete or partial interviews. We use data from CATI survey conducted in June 2016 by Public Opinion Foundation in Saratov region, Russia. 7210 interviews have been completed. We determine questions that are likely to cause undesired interviewer behavior such as questions with long answer options list. Then interviews eligible for control are identified both by random assignment and picking cases in which time to complete question noticeably differs from average. 1125 complete interviews were monitored totally. We measure frequencies of interviews flagged by monitors as conducted with multiple violations of instruction. Remarkable differences in these frequencies are discovered between randomly selected interviews and those marked as suspicious with our method. Up to 90% of non-quality interviewer work are found in flagged interviews that amount to 10% of all cases. Application of this method allows researcher to eliminate poor-quality cases with higher precision and reduced costs and also to provide interviewers with grounded feedback. While the case examined is telephone survey we believe it can be applied to computer-assisted paper interviews as well.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Methodological Brief: Interviewers, Quality Control & Sample Design
A Probability Based Sample of Family Planning Clinic Patients in Delaware
Michel Boudreaux *University of Maryland*
Michael Rendall *University of Maryland*
Steven Martin *University of Delaware*

There is an extensive literature examining the use of family planning services based largely on convenience samples of clinic patients that might not generalize to larger populations. In this paper we describe and evaluate a novel probability based approach. The target population consisted of female, reproductive age clients of Title X clinics in the state of Delaware. The frame consisted of a complete list of Title X clinics, a list of operating hours, and tabulations of clinic specific family planning user populations. These tabulations were generated from the clinics' administrative records for federally mandated Family Planning Activity Reports (FPAR). Using these sources we drew a multi-stage, stratified and clustered sample of patients. The 21 eligible clinics were stratified based on population size. Large clinics were sampled with certainty and smaller clinics were systematically sampled. We collected data on 500 clients allocated across sites according to the distribution of patients implied by the FPAR. To ensure that all patients had a chance of being interviewed at any time during the day and on any day during the fielding period, we randomly ordered the sequence in which the field staff visited the sampled clinics. When the field staff visited a clinic they interviewed all women that showed up for care, from opening to closing. This approach to time sampling had both efficiency advantages and abided by clinic rules which stipulated that all patients present during fielding days had to be given the chance to participate (i.e. there could be no within clinic sampling of patients). Data collection will end in early 2017. We will assess the representativeness of the sample by comparing the demographics of the sample to demographic distributions in the FPAR. We will examine how our approach compares to a hypothetical convenience sample drawn by visiting a single clinic.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Methodological Brief: Interviewers, Quality Control & Sample Design

How to Break Into Jail: Practical Considerations for Interviewing Incarcerated Respondents

Jodi Walton *Abt Associates*

Donna J. DeMarco *Abt Associates*

Jo Anna Hunter *mdrc*

Increasing policy and media attention on criminal justice reform requires rigorous, high quality evaluation of the social programs that serve disadvantaged populations and those vulnerable to criminal activity. This research is critical for identifying effective interventions. The U.S. Department of Labor's Enhanced Transitional Jobs Demonstration (ETJD) is designed to serve two such populations: ex-offenders and non-custodial parents. The ETJD offers job training designed to build skills needed to compete in today's workforce. Members of these vulnerable populations are notoriously difficult to survey: transience and incarceration pose significant challenges to locating and interviewing. In this brief, we describe our efforts to locate and interview 5,000 ETJD participants, 77 percent of whom were ex-offenders, at 12 and 30 months post participation in the ETJD program intervention. We show that despite increased rates of incarceration among sample members in the 30 month follow-up, survey response rates improved from 75% (12 month) to 77% (30 month). We hypothesize that these rates resulted from increased access to state and county facilities in the 30-month follow up. In this brief we describe the most successful tools for conducting respondent interviews in state and county prison facilities that results in efficient, effective research outcomes. We will describe protocol design issues essential to gaining IRB approval; securing access with state and county departments of justice/corrections and interviewer training and preparation. Many of the considerations that will be discussed would also be applicable to other institutionalized populations, such mental health or long term care facilities.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Methodological Brief: Interviewers, Quality Control & Sample Design

Developing and evaluating a gradation assessment index for survey data quality assurance practices

Y. Patrick Hsieh *RTI International*

Joe Murphy *RTI International*

Annice Kim *RTI International*

Jamie Guillory *RTI International*

Brian Bradfield *RTI International*

Satisficing theory suggests a battery of measures researchers can use for assessing the extent to which respondents provide high quality responses. Individually, these measures can be helpful in pinpointing respondents who put forth less effort than was demanded by the question or questions at hand. However, researchers may encounter significant practical challenges applying satisficing theory as the availability of individual quality indicators mainly depends on the design of the survey instrument. Moreover, the effectiveness of a given indicator may also depend on the contextual factors such as the placement in questionnaire and the substantive focus of research. To address such practical challenges in assessing survey data quality, we developed a gradation index that integrates multiple measures of satisficing into a single index: item-nonresponse, attention filters, response consistency indicators, and non-differentiation. We then employed the scheme to a web survey measuring respondents' usage of e-cigarettes and their exposures to commercials or information about e-cigarettes from mass and social media. To gauge the potential impacts of satisficing by respondents on the substantive estimates, we examined estimates using different thresholds of the satisficing gradation index. By ranking respondents by their overall level of satisficing, we were able to assess the impact of various cut points of the index for data quality on the substantive survey estimates. We find that the respondents scoring higher on the satisficing index tended to report more frequent usage of e-cigarettes and higher exposure of e-cigarette information. We discuss the potential effect of questionnaire design and respondent behaviors that might lead to such response patterns. We conclude our study by discussing the considerations for surveys to prioritize of the data quality measures given the instrument design and the contexts of research questions.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Methodological Brief: Interviewers, Quality Control & Sample Design
River Sampling and Underrepresented Voters: Addressing the Challenges of Efficiency and Security
Matthew Towery *Opinion Savvy, LLC*

During the 2016 election cycle, the field of public opinion polling was dealt several challenges: the uniqueness of the presidential candidates and accompanying news cycles; the decline in response rates for traditional methodologies; and the increasingly difficult prospect of contacting millennials, ethnic minorities, and disengaged voters. This brief will examine the results of my own company's publicly released polls for both the 2016 presidential primary and the 2016 general election at the statewide level. In particular, the brief will examine the relative effectiveness of reaching younger voters and ethnic minorities, as well as the "shy" voter phenomenon that has received widespread media attention. The brief will also address some of the challenges facing this methodology, in particular the issue of safeguarding the sample from outside intervention, as well as the process of matching voter records to potential online respondents.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Making American Great: Domestic Policy

Time, Space, and Attitudes toward US–Mexico Border Security

Timothy B. Gravelle *Wilfrid Laurier University*

The tumultuous 2016 US Presidential Election cycle featured a range of policy proposals to address the issue of illegal immigration. Channeling anxieties around the economic and social consequences of illegal immigration with claims of porous, unsecured borders, Republican candidate Donald Trump notably committed to building a wall the length of the US–Mexico border. At the same time, border security is not a new issue on the policy agenda, and its salience has risen and fallen over time. Drawing on multiple surveys over the period 2006 to 2016 and spatial analytic tools, this paper explores two questions. First, how have attitudes toward border security shifted over time in response to changes in the partisan political environment? Second, how does spatial context – namely proximity to the US–Mexico border – shape attitudes toward the proposed border wall? Findings point to both time and space, in conjunction with individual-level political factors, as key factors shaping attitudes toward US–Mexico border security.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

We Miss You. A Little Too Much, A Little Too Often: Nonresponse Prevention, Evaluation and Adjustment

Selection Sensitive Survey Design

Michael Bailey *Georgetown University*

Non-response is a large and growing problem in survey research. It is well known that weighting, the most widely used tool for dealing with non-response, exacerbates non-response bias when there is selection on unobservables. Nonetheless, there is surprisingly little effort in the literature to diagnose whether such endogenous selection occurs. One reason for this is that most survey data sets lack the type of information needed for tests of endogenous selection to perform well. This paper presents two tools that facilitate diagnosing and correcting for endogenous selection. First, one can use randomized opt-in questions in order to generate a variable that explains response, but does not affect outcome variables directly. Second, after eliciting willingness to respond, one can follow up with political questions for everyone. One can then easily assess weighting models' assumption that willingness to respond is unrelated to opinions. Empirical examples from two surveys demonstrate that these tools are easily implemented and provide information that enables effective diagnosis of endogenous selection.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

The Election, Polls and the Media

Perceptions of Polls and Voter Expectations: Competitive Poll Results, Methodology, and Opinionation

Ozan Kuru *University of Michigan*

Josh Pasek *University of Michigan*

Michael Traugott *University of Michigan*

Although we know a lot about how people interpret individual poll results in news stories, little systematic work has examined what happens when people encounter conflicting poll results. The inconsistency in poll results would not always reflect real shift in support levels but could rather be indicative of the differences in methodologies. Although the general public might not be very sensitive to methodological issues, experts and partisans heavily weigh in to provide commentary on poll results that might be objective or not. How do people perceive and react to the polls in such a competitive information environment where polls' results, methodological qualities, and their interpretations vary substantially? Whereas a substantial body of research has examined how horse-race coverage leads to changes in public skepticism and voting, no previous study tapped these dynamic and opinionated coverage of polls. Drawing from the motivated reasoning, corrective attempts, and opinionated news literatures, we test the impact of competing poll results (favoring either candidate), competing methodological quality of polls (robust-vs-poor), and competing commentaries on polls (how experts and partisans debunk or praise polls) on people's perceptions of polls and candidate performance in the context of the 2016 U.S. presidential election. On a nationally representative sample of Americans (Time-Sharing Experiments in Social Sciences, GfK, N=2078), we conducted a cross-sectional survey experiment in which we manipulated these factors and measured respondents' perceptions of poll reports as well as their predictions of the support that each candidate would get. Our results show how motivational biases are triggered, amplified and could be mitigated in this dynamic, competitive, and opinionated information environment. We discuss the implications for the nature of public opinion in the context of horserace politics and motivational biases.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Questionnaire Design in the 3MC Context

A Test of Generalization of Classic Question Order Effects in Different Cultures

Tobias Stark *Utrecht University*

Jon Krosnick *Stanford University*

Henning Silber *GESIS - Leibniz-Institute for the Social Sciences*

Annelies Blom *University of Mannheim*

This research tested whether findings of question order experiments, reported in the U.S. three decades ago, would replicate in U.S. online surveys today and whether these findings would generalize to 11 other countries (Canada, Denmark, Germany, Iceland, Japan, the Netherlands, Norway, Portugal, Sweden, Taiwan, and the United Kingdom, total N = 25,607). One question order effect involved the norm of evenhandedness asking about financial contributions of businesses and unions to political campaigns and the other involved a perceptual contrast between two questions about abortion. We proposed that question order effects due to the norm of evenhandedness only occur if respondents prefer one of the entities the questions are about to the other entity (here businesses or unions). This condition was met in the U.S. and in 6 other countries. The question order effect replicated in the U.S. and generalized to all but one country (Iceland) in which the necessary condition was met. We further proposed that the question order effect due to perceptual contrast should only occur if respondents consider one of the two question to give a much more compelling reason for abortion than the other. This condition was met in all but the three Scandinavian countries in our sample. Also this question order effect replicated in the U.S. Interestingly, it generalized to all countries, not just the ones that met the necessary condition. This finding suggests that varying cultural norms between countries can make it difficult for researchers to detect a response effect caused by a perceptual contrast.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

We Miss You. A Little Too Much, A Little Too Often: Nonresponse Prevention, Evaluation and Adjustment

Managing Respondent Burden for a Household Panel using Permanent Random Number Sampling

Nadarajasundaram Ganesh *NORC at the University of Chicago*

Vicki Pineau *NORC at the University of Chicago*

J Michael Dennis *NORC at the University of Chicago*

Multi-client survey platforms such as probability or non-probability household panels may inadvertently place a disproportionate respondent burden on a select number of panelists, increasing the risk of panel attrition and panel conditioning by over-engaging some panel segments and under-engaging other segments. Unless statistical controls are in place, some panelists might be sampled more frequently for client studies, while others potentially might rarely be sampled for client studies. In this presentation, we discuss the mechanics and impacts on panel management by using permanent random number sampling for selecting panelists for client studies conducted using the AmeriSpeak® Panel, a probability-based, mixed-mode household panel operated by NORC. The use of permanent random number sampling overcomes the problem of unequal respondent burden by “equally distributing” client survey assignments across the entire panel. We present results comparing respondent burden among panelists when using permanent random number sampling against simple random sampling that does not attempt to mitigate overlap in the selected sample across different client surveys. Furthermore, we discuss our approaches to modifying permanent random number sampling to select at most one panelist per household for a given client survey and minimizing unequal weighting effect due to nonresponse to the client study.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Fit Your Purpose: Frameworks and Examples of Alternatives to Probability Sampling

Differences in Probability and Non-probability Samples of Asian Racial Subgroups

Bryan B Rhodes *RTI International*

Ellen L Marks *RTI International*

Darrick Hamilton *The New School*

William A Darity *Duke University*

When surveying rare populations, researchers often must weigh the analytic needs of obtaining a representative sample against the high costs of identifying members of these populations. Nonprobability methods allow researchers to more easily identify and target rare populations compared with probability samples, thereby reducing survey costs; but nonprobability samples may introduce bias. While previous studies have compared assets and debts across broad racial group levels (e.g., blacks, Asians or Hispanics), the National Assets Scorecard for Communities of Color (NASCC) looks more closely at differences in assets and debts among subgroups within these categories (e.g., Salvadorans, Hondurans or Mexicans). Because these subgroups typically represent a small portion of the total U.S. population, NASCC has focused on specific metropolitan areas, like Los Angeles, where specific subgroups of interest are more commonly found. In the most recent NASCC study, in-person interviews were conducted in Los Angeles using an address-based probability sample with flags to indicate the household's race/ethnicity. However, this method did not achieve sufficient sample sizes among two of the hardest to locate subgroups, Koreans and Cambodians. As a result, we conducted intercept surveys at known gathering places for these populations. This paper will first outline the methods used for both the probability and non-probability based sampling. We will then compare key survey estimates such as age, income, marital status and other demographic characteristics for respondents who completed an interview from the probability-based sample and those who completed from the non-probability based sample. Examining these differences can help identify areas of potential bias in the non-probability sample or non-comparability across the two samples. We will also explore the relative costs of each method.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

We Miss You. A Little Too Much, A Little Too Often: Nonresponse Prevention, Evaluation and Adjustment

Evaluating Selection Bias in a Multi-Phase Health Survey

Caroline Blanton Scruggs *RTI International*

Marcus Berzofsky *RTI International*

Bo Lu *Ohio State University*

Timothy Sahr *The Ohio Colleges of Medicine Government Resource Center*

In surveys there is a chance that the distribution of sampled respondents will not match a target population, which can result in a less representative survey and reduced precision in estimates. In multiple phase studies the sample is confined to respondents from a prior stage of the study. The decision to continue to later phases may be influenced by the respondent's interest in the study, availability or factors that aren't measurable in earlier phases. If this occurs then there is the potential for selection bias in later phases (i.e. self-selection into the second phase rather than a random sample). The Ohio Medicaid Group VIII (G-VIII) Assessment is a study of Ohio's Medicaid population with the goal of understanding how Medicaid participants who were enrolled via the state's Affordable Care Act associated Medicaid expansion compared to those who were Medicaid enrolled prior to expansion. To make the comparison the G-VIII had four main phases: (1) a telephone survey, (2) a biometric screening, (3) a medical records extraction; and (4) Medicaid administrative data analyses. The phase 1 telephone survey consisted of a stratified random sample of Medicaid enrollees. All persons who completed the telephone survey were asked to participate in a biometric screening and all persons who participated in the biometric screening were asked to participate in medical records extractions – Medicaid claims analyses were used to profile the sample universe. The biometric screening had an additional level of nonresponse among those who agreed to participate in the screening but did not appear at the prescribed appointment time. This paper examines G-VIII nonresponse to assess if selection bias may have occurred through comparison of respondent distributions and weighted distributions. We discuss strategies to reduce the impact of any potential selection bias through multiple weighting adjustments to correct for each phase of nonresponse.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Measurement Effects in Surveys

Improving Tests for Straight-Lining

Benjamin Phillips *Abt Associates*

Valrie Horton *Abt Associates*

Jared Knott *Abt Associates*

Rosaella Branson *Boys and Girls Clubs of America*

Elizabeth Fowlkes *Boys and Girls Clubs of America*

Straight-lining (i.e., identical responses to each row of a grid item) is a recognized data quality problem in web surveys. Respondents who straight-line are seen as satisficing, calling into question the veracity of their responses. Commonly used approaches set a threshold for determining whether a given respondent is problematic based on the proportion of grids the respondent has straightlined (e.g., 25%). This approach has potential drawbacks, as not all grids are created equal. The probability of a nonsatisficing respondent straightlining is higher on grids with fewer rows. In addition, some grids may simply have little variance (e.g., a satisfaction scale for popular institutions) and the probability of nonsatisficing respondent straight-lining will be higher here, too. A simple threshold approach based on the percentage of grids straight-lined risks falsely identifying some nonsatisficing respondents as satisficers. Removing false positives from the sample will decrease yields and bias estimates. We present two developments that reduce the likelihood of false positives. The first applies inferential statistics to test whether the observed distribution of straight-lining in a grid differs from the amount of straight-lining that would be expected given the distribution of responses. This provides a means to determine whether satisficing-motivated straight-lining is a concern. The second is a scoring system for straight-lining across a survey that accounts for the likelihood of nonsatisficing straight-lining. Straight-lining on grids where a higher degree of straight-lining would be expected given the distribution of responses is scored lower than is straight-lining where straight-lining is less likely. This approach reduces the risk of false positives. Summed scores above a given level (e.g., the 99th percentile) can be removed from the sample or, better, subjected to additional scrutiny for other indicators of satisficing, such as short length of survey. These approaches are demonstrated on a very large national youth survey.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Questionnaire Design in the 3MC Context

Questionnaire Design in Short Message Service (SMS) Surveys: Split Ballot Experiments in Four African Countries

Herschel Lisette Sanders *RTI International*

Charles Lau *RTI International*

Ansie Lombaard *Kantar*

Collecting data via short message service (SMS, or text messaging) is growing in popularity. Use of SMS is particularly common in lower income countries, where mobile phone penetration is increasing and face-to-face surveys are particularly challenging. SMS offers a low cost, rapid, and highly accessible mode of data collection given the ubiquitous use of cell phones. However, writing questions for SMS surveys can be difficult. Space is limited; most respondents use older phones with small screens; questionnaire designers only have 160 characters to write both the questions and response options, and SMS surveys can only accommodate a few questions. And, respondents may have limited literacy. Although there has been some research on SMS surveys, the focus has primarily been on representativeness and their use in asking sensitive items. This study will further the knowledge in best practices for questionnaire design within the SMS survey context. We study questionnaire design in SMS surveys using survey data with experimental designs from Ghana, Kenya, Nigeria, and Uganda (n = 9,697). The analysis investigates three specific questions that SMS questionnaire designers commonly face. First, is it feasible to ask “select all that apply” questions on SMS surveys? Second, should SMS questionnaires include a “don’t know” response option or not? Third, does the order of response categories affect how respondents answer questions? This paper will evaluate the three experiments using descriptive analyses. The results highlight difficulties with “select all that apply” questions and “don’t know” responses that may impact both the understanding of the survey and resulting data quality. We also explore response order effects and how they appear to operate in this unique survey response environment. Finally, we discuss the practical implications of the study for researchers who wish to use SMS as a data collection platform for surveys.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

We Miss You. A Little Too Much, A Little Too Often: Nonresponse Prevention, Evaluation and Adjustment

Comparison of weighting procedures in the presence of unit nonresponse: a simulation study based on data from the American Time Use Survey

Morgan Earp *Bureau of Labor Statistics*

David Haziza *University of Montreal*

Nonresponse has the potential to induce biases in the estimates, if units who do not respond to the survey are systematically different from those who do. The most common way to deal with unit nonresponse is through a weight adjustment procedure. This procedure typically consists of eliminating the nonrespondents from the data file and adjusting the weights of the respondents, with the goal of reducing nonresponse bias. We distinguish between two types of weighting procedures: in the first base weights are multiplied by the inverse of the estimated response probabilities, whereas the second uses some form of calibration to adjust the base weight. Commonly used methods for estimating the response probabilities include logistic regression, weighting classes based on logistic regression and regression trees. Commonly used calibration procedures include linear, exponential and truncated linear weighting. This paper summarizes a study which compares these two weighting procedures in terms bias and efficiency using a simulation study based on data from the American Time Use Survey.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Fit Your Purpose: Frameworks and Examples of Alternatives to Probability Sampling

Surveying Rare or Hidden Populations Using a Probability-Based Household Panel

Vicki Pineau *NORC at the Univ of Chicago*

J. MichaelDennis *NORC at the University of Chicago*

Stuart Michaels *NORC at the University of Chicago*

Sherry Emery *NORC at the University of Chicago*

Nadarajasundaram Ganesh *NORC at the University of Chicago*

Estimating economic, health, and social disparities among priority subpopulations (e.g., LGBT) is increasingly regarded as essential for policy making and scientific inquiry but is problematic without resorting to non-probability sampling. Traditional probability-based sampling strategies are impractically expensive because of the large scale of the in-field screening required to find sufficient numbers of persons in so-called “rare” or “hidden” populations. As a result, rare or hidden populations are often studied using less rigorous methods such as “snow-ball” sampling and non-probability, opt-in web panels. Our paper is based on an NORC pilot study testing a cost-effective alternative for surveys of rare or hidden populations. The tested approach combines probability sampling and Network and Respondent-Driven Sampling (RDS). The target subpopulations for the pilot study are the lesbian, gay, bisexual, and transgender communities. The authors will present the pilot study findings with respect to: (1) assessing the feasibility of using a probability-based panel sample, with multiple rounds of nominations of non-AmeriSpeak panelists by AmeriSpeak panelists, to survey a larger sample of people who self-identify as L, G, B, or T in the U.S; and (2) comparing estimates of key demographic and health outcome estimates from the survey under a multiplicity counting approach for Network Sampling and an approach based on RDS. The initial sample source for the pilot study is NORC’s AmeriSpeak Panel®, which is an online household, multi-client panel that uses the probability-based NORC National Frame to construct an address-based nationally representative sample panel with sample coverage of approximately 97% of US households.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Making American Great: Domestic Policy

How Racial Attitudes Affect Public Opinion on the Economy: Assessing Blame and Credit for Economic Change

David C Wilson *University of Delaware*

Darren Davis *University of Notre Dame*

We use survey experimental data from the Cooperative Congressional Election Study (CCES) to consider the extent to which national (Obama) and state (Governors) leaders are held responsible for reported changes in national and state economies. We asked respondents how much responsibility political leaders have for changes national and state economies, randomizing the description of change as “improves,” “worsens,” or “changes.” Results show that the direction of change only matters for the national economy. Political leaders tend to get more responsibility when the national economy “changes” or it “worsens;” lower levels of responsibility are given when the economy improves. These findings suggest that baseline considerations for national economic change in 2012 were that it would worsen. There were no significant differences in perceptions of responsibility for state economic change. Perhaps, the most important finding is that the direction of the results are predicted by racial resentment.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Innovations in Cognitive Interviewing: Recruitment Strategies and Interviewing Techniques

Conducting Cognitive Interviews with Young Children

Rachel Levenstein *American Institutes for Research*

Rachel Levenstein *American Institutes for Research*

Mahlet W. Megra *American Institutes for Research*

Jen Durow *American Institutes for Research*

Evan Nielsen *American Institutes for Research*

As a result of the Every Student Succeeds Act, education policy at the federal or state level may require schools and districts to survey children as young as second grade (aged 7-8) about school climate or classroom practices. However, many methodologists urge caution in surveying young children. In a review of the literature, Borgers et al. (2000) discourage surveying children from ages 4 to 8 because they may have limited language abilities, interpret questions quite literally, be more likely to respond in a way to please adults, and have shorter attention spans compared to older children or adults. If younger students will be surveyed as required by federal or state policy, existing surveys that are used for older students may need to be adapted by shortening the survey and simplifying the language. Cognitive interviews may be useful in this adaptation process. In such interviews, participants are asked to read survey questions and think aloud as they respond, reporting to an interviewer the cognitive processes they use in answering the questions. These reports can easily be used to improve survey questions. However, young children may not be developmentally able to explain how they arrived at a particular response. To date, no research has examined the utility of conducting cognitive interviews with young children. This paper will examine the effectiveness of cognitive interviews with 2nd-8th grade students in eliciting meaningful feedback that can be used to improve survey questions. The sample will include 35 students who were administered surveys about classroom practices and personalized learning in the fall of 2015 and the fall of 2016. Using qualitative analyses, we will examine interviewer probes and think-aloud responses to understand if younger children can self-report on their cognitive processes and if particular probes are more or less effective at eliciting useful feedback.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Innovations in Cognitive Interviewing: Recruitment Strategies and Interviewing Techniques

Remote Cognitive Interviewing: An alternative to traditional in-person cognitive testing

Aleia Clark Fobia *U.S. Census Bureau - Washington, DC*

Jennifer Childs *U.S. Census Bureau*

Cognitive interviews are traditionally conducted in person. However, remote cognitive interviewing – where the interviewer and respondent are not in the same location— provides several advantages over traditional methods of in-person cognitive interviewing. In-person cognitive interviews are expensive since they require an interviewer and respondent to meet in a common location. Respondents are also paid an incentive and the number of interviews is usually low. Remote testing over the web can significantly increase the number of respondents in a cognitive pretest. It sidesteps the problem of geographic constraint, and paying incentives may not be necessary. However, questions remain about best practices for remote cognitive interviewing. In this presentation, we report the results of three separate cognitive projects. Each project included both remote and in-person cognitive interviews. This side-by-side design provides a unique opportunity to analyze differences in data from remote and in-person cognitive interviews. We explore differences in responses and respondent characteristics, which probes work best in remote testing, and advantages and limitations to remote cognitive interviewing. We offer suggestions for further research on how best to combine in-person and remote testing. Through direct comparison with more traditional methods, this research will contribute to literature on alternatives to traditional in-person cognitive interviewing.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Measurement Effects in Surveys

Mode Effects within the Same Individual between Web and Mail Administration

John Boyle *ICF International*

Robert Tortora *ICF*

Bryan Higgins *ICF*

Naomi Freedner *ICF*

Mode effects are often assumed when survey estimates for the same measure vary significantly by wave in multi-mode surveys. However, these differences may also reflect differences in the populations who respond by mode in these surveys. We rarely have test/retest data from large sample with the same individuals to the same measures in different modes that would verify real mode effects. As a result of a coding error in the implementation of a national ABS survey, a sample of approximately 4,600 respondents who initially responded by web were subsequently sent a request to complete a paper questionnaire. A total of more than 600 unique respondents with the same age and gender at the same address completed the same interview by mail. There were two versions of the questionnaire, which included attitude, behavior and demographic items. Version A included a total of 46 items for all respondents, and an additional 75 items for some respondents. Version B included a total of 36 items for all respondents, and an additional 135 items for some respondents. Some questions were shared by both versions of the questionnaire while others were unique to the specific version. In order to estimate mode effects, we perform a multitrait-multimethod matrix analysis on key measures in the survey to evaluate convergent and discriminant validity as well as to estimate method variance.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Measurement Effects in Surveys

An Assessment of Bias in Estimates using Data from the NHIS Self-reported Web Users

Meena Khare *NCHS/CDC*

Nonresponse and increasing survey costs continue to threaten traditional methods of data collection. In recent years, alternative data collection methods have been explored by survey practitioners and researchers. According to the Pew Research Center telephone surveys, the rise in Internet users has grown from 14% of the U.S. adult population in 1996 to 84% in 2015. Therefore, Web surveys and multimode data collection from households and establishments have become a common, alternative cost-saving approach. The National Health Interview Survey (NHIS) has been collecting information on the Internet and email usage among adults since 2012. Among 2014-2015 NHIS adult respondents, 74.1 % report using the Internet or emails. This paper presents an assessment of bias in selected estimates by comparing estimates computed from a random sample of self-reported Web users (defined as NHIS adult respondents who use the Internet or email) with the overall NHIS estimates.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Measurement Effects in Surveys

Measuring happiness and life satisfaction amongst Swedish citizens: An inquiry into semantic equivalence in comparative survey research

Sofia Sigrid Maria Axelsson *Department of Political Science, University of Gothenburg*

Stefan Dahlberg *Department of Political Science, University of Gothenburg*

This paper addresses translational effects in comparative survey literature, focusing on measures of happiness and life satisfaction which, in recent years, have become increasingly central to studies of the contemporary political and social world. Measures of self-reported happiness and life satisfaction are frequently applied by large-scale comparative surveys and presented as international indices where Scandinavian countries generally occupy top positions. Cross-national variations accounted for by survey data are often explained to be the result of political, economic and social factors such as level of development or form and content of government structures. Even as the importance of language and culture are frequently referred to in the literature, translational effects are often treated as a secondary concern by the quantitative research community. In an effort to contribute to the debate, this paper presents unique findings from survey experiments with a large sample of Swedish citizens. Inspired by an experimental design applied in Denmark, we set out to explore differences in question wordings and response scales, randomizing respondents into Swedish or English language groups. The results reveal significant differences in self-reported levels of happiness between Swedish and English questions. While the mean differences are small, the distribution of answers between language groups is substantial enough to confirm a semantic gap between the English term "happy" and its Swedish translation "lycklig". Hence, it requires something more to be "very happy" in Swedish than in English. Interestingly, language appears to have a lesser impact on the distribution of responses across language groups when using a numbered response scale, indicating that a particular question design either mitigates or intensifies translational effects. Happiness, it is concluded, is not easily translated and survey practitioners should bear this caveat in mind when operationalizing the concept across countries and cultures.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Measurement Effects in Surveys

Patterns in panel effects: A meta-analysis.

Andrew Caporaso *Westat*

Rosalynn Yang *Westat*

Jingwei Hu *JPSM*

Roger Tourangeau *Westat*

Longitudinal data collection is an important method for measuring change in a population over time. A concern with longitudinal research is the possibility that simply participating in the research can affect the behaviors and responses of study participants, a phenomenon referred to as ‘panel conditioning’ or ‘panel effects.’ If the actual behaviors or answers of respondents are influenced by the study itself, then the validity of inferences made by longitudinal studies could be called into question. Dozens of studies have used experimental methods to quantify the impact of panel effects in survey research, finding mixed results regarding the presence, direction, and magnitude of such effects. Common explanations for panel effects include the ‘cognitive stimulus hypothesis’ which suggests that most people do not have well-formed opinions on most topics and being asked to report opinions forces them to develop opinions over time (Sturgis, Allum, Brunton-Smith, 2009). Another explanation is that respondents learn how to circumvent the interview by answering questions in a way that reduces burden (Toepoel, Das, and van Soest, 2008). The goal of this report is to review and synthesize over 30 reports on panel effects in health, economic and social research. It will attempt to meta-analyze the experimental data on panel effects to determine whether there are consistent patterns in the size and direction of such effects in economic, health and social research. The findings will be useful for researchers to better understand and predict how panel effects might influence their particular studies.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

The Election, Polls and the Media

“Gauging Public Opinion in the Age of Trump”

Robert Martin Eisinger *Roger Williams University*

Abstract: Gauging Public Opinion in the Age of Trump Election night 2016 shocked pollsters, even many of the Republican ones, not to mention the vast majority of the media pundits. Nate Silver’s commonly read and reputable prognostication/prediction website, 538, showed a 71.4% probability that Hillary Clinton would win the presidency on Election Day. For academics, who disproportionately did not vote for Donald Trump, the results on Tuesday night and Wednesday morning were beyond disappointing. The zeitgeist of the nation – their zeitgeist – had been disrupted and dislocated. The world as they knew it was no longer. Perhaps it never was. This paper seeks to commence and reignite discussions about what constitutes public opinion, and the challenges of measuring it. Reviewing a literature from the 1880s to the present, I will discuss how conceptions of public opinion often include group dynamics and political apathy/ignorance – concepts that public opinion scholars and practitioners all too often overlook or ignore. While the discussions about sampling error are valuable and important to advance the polling techniques, public opinion practitioners and scholars would benefit from rekindling dialogues about the construction and formation of public opinion, as they will further aid us in understanding what exactly we are seeking to measure, document and analyze. Only if we recognize that some expressed opinions are amorphous and inconsistent, can we then advance conversations about what roles public opinion should play in democratic governance.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Innovations in Cognitive Interviewing: Recruitment Strategies and Interviewing Techniques

Using Online Panels to Approximate Populations of Interest for Cognitive Testing

Jessica L. Holzberg *U.S. Census Bureau*

Lucia C. Lykke *U.S. Census Bureau*

Previous research has found that conducting online, unmoderated cognitive interviewing using services such as Amazon Mechanical Turk have potential as a component of a broader pretesting technique (e.g., Murphy, Keating, & Edgar 2013; Edgar 2013; Fowler et al., 2015; Cook et al., 2015). Online cognitive testing, or web probing, is typically conducted with the general population, but testing some surveys and forms requires recruiting respondents who meet specific criteria such as having advanced degrees, complex households, or children receiving free school meals. When these more specific populations are needed, researchers currently screen within the online instrument or recruit elsewhere (e.g., Craigslist) then direct respondents to the Web. These methods are not always feasible. Alternatively, using an online panel as a recruitment source allows researchers to preemptively restrict online cognitive interview respondents based on demographic characteristics. In this paper, we suggest that using an online panel may be appropriate when sampling a very specific population of interest: families with children receiving free and reduced price lunches. Although this particular recruitment criterion cannot be stipulated in advance using an online panel, we know the demographic characteristics of this population. Conducting online cognitive interviewing with members of a similar population may provide an inexpensive method of cognitive testing that complements more traditional methods. We present results from concurrent cognitive interviewing efforts: traditional in-person cognitive interviews with adults who applied for free or reduced price meals for household children, and online interviews with a panel whose demographic characteristics mirrored those of the in-person respondents. We compare responses to similar probes between the two groups, and administer additional probes online that could not be included in person due to time restrictions. We discuss the usefulness of testing materials online with a population that demographically approximates a specific population of interest.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

The Election, Polls and the Media

Candidate Personalities and Political Issues: A Content Analysis of Major Public Opinion Issues in Traditional News Media during the 2016 U.S. Presidential Election

Thomas B Christie *Univ. of Texas Arlington*

This study examines candidate and issue information in traditional news media during the 2016 U.S. Presidential campaign between Hillary Clinton and Donald Trump, useful in understanding media agenda setting and the possible influences on public opinion in the key stages of a major presidential general election. The paper examines the 2016 period of intense national media coverage of issues drawn from leading public opinion polls. This study builds upon an earlier study finding that even though leading sources of news may have provided proportionally less coverage of issue information in favor of candidate information at the beginning of a presidential primary, the amount of issue information provided to newspaper readers increased toward the end of the campaign. Key research questions informing this study were: 1. What were the most prominent issue and candidate information categories (from public opinion polls) covered by leading national newspapers during the 2016 general election? 2. What was the frequency of coverage of issue and candidate information during the campaign by leading newspapers? 3. How did coverage of key exit poll issues compare to coverage of other categories of issue information? 4. Were there differences between issue and candidate information coverage among the newspapers during the course of the campaign? The timeframe of this study begins with the date the presumptive nominees were established--June 6, 2016--and ends with the date of the general election, Nov. 8, 2016, allowing for an analysis of prominent issue and candidate information influencing public opinion up to the date of the general election. A detailed analysis of the coverage of these issues compares them to the public's ranking of issues in major public opinion polls, including exit polling, and reveals major changes in how these issues were covered over the course of the presidential campaign.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Making American Great: Domestic Policy

Trumped on Trade: Navigating Perceptual Barriers Between US Elites and the Public

David Rankin *State University of New York at Fredonia*

For decades, national surveys have consistently reported an opinion divide between American leaders/elite and the mass public on the topic of trade liberalization and related agreements. In the 2016 presidential elections, these divisions came to an electoral head with the upstart Democratic Sanders and Republican Trump campaigns. Several studies have concluded higher educated and skilled US “elite” to be more likely to favor policies of trade liberalization than the average American based on economic self-interest, and at times of more salient trade debate, influenced by predispositions such as ideology and partisanship. Fewer studies have examined the heuristic impact of affective attitudes and symbolic predispositions, such as enemy images and national identity, particularly in a comparison of American leaders and the mass public when it comes to views on trade. In this paper, I examine and compare how American leaders/elite and citizens differ in their policy judgment on trade liberalization and agreements, including NAFTA and the TPP. I analyze the relative impact of occupational skills, education, geographic location, partisanship and ideology, while exploring how symbolic predispositions, namely conceptions of national identity, significantly impact trade policy judgment across the American public. The analysis focuses on the time period, 1990-2016, utilizing national survey data drawn from the National Identity modules of the International Social Survey Programme, and the quadrennial surveys of US leaders/elite and the general public conducted by the Chicago Council on Foreign Relations/Global Affairs and the Pew Research Center.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Making American Great: Domestic Policy

Small Business' Reactions to State Sponsored Retirement Security Policies

Andrew Blevins *The Pew Charitable Trusts*

Theron Guzoto *The Pew Charitable Trusts*

John Scott *The Pew Charitable Trusts*

Alison Shelton *The Pew Charitable Trusts*

Sarah Spell *The Pew Charitable Trusts*

Thomas Brassell *ICF International*

Robert Tortora *ICF International*

James Dayton *ICF International*

Brian Orleans *ICF International - Fairfax, VA*

Though most Americans will rely on personal savings to supplement Social Security benefits in retirement, one-third of private sector employees lack access to an employer sponsored retirement plan. State policymakers are actively exploring policies aimed to increase access to private sector retirement plans, including enabling employers to begin their own plans, such as through marketplaces or multiple employer plans, or to enroll workers in auto-IRA plans sponsored by a state or city for workers without plans at their employer. To create effective policies, understanding the needs and concerns of small businesses is needed. Using data are from a nationally representative telephone survey of small, private-sector businesses, we examine how employers react to retirement policy initiatives, including auto-IRA plans, multiple employer plans, and marketplace exchanges. Which policies are most supported? How do establishment characteristics relate to support and opposition to such policies? Focus group research suggested that employers with experience with pro savings features like automatic enrollment, automatic escalation, and employer contributions would view auto-IRA programs more favorably than those who did not. Additionally, preliminary focus group research had indicated that employers might not find a retirement plan marketplace useful. Preliminary analyses show businesses generally support these policies. Employers with plans that include pro-savings features were more likely to somewhat support auto-IRA plans than oppose them. Counter to initial focus group research, a significant majority of employers said that a marketplace would be helpful but only a modest majority said it would make them more likely to offer a plan. Those who believed they would be somewhat or very likely to start their own plan soon were less supportive of multiple employer plans. This paper provides guidance for policymakers searching for the best way to help their constituents.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

The Election, Polls and the Media

Diminished Voices: Polling, the Press, and the Representation of Minority Perspectives in Political Discourse

Shakari Byerly *University of California, Los Angeles (UCLA)*

This paper will explore the use of polling by print news outlets and the representation of minority interests in their reporting of mass opinions. Pioneers of polling such as George Gallup, Elmo Roper, and Hadley Cantril envisioned polls as a mechanism for conveying the popular will. However, the emphasis on state and national polls shifts the focus towards majority populations. This is in contrast to the Madisonian vision of democracy, which emphasizes the importance of the opinions of political minorities. Much of the research on the role of polling in the democratic process focuses on how public opinion is measured, whether it can be measured, how the publication of polling results impacts electoral outcomes, or how it is utilized by interest groups and political campaigns to manipulate public opinion and advance the self-interest of the entities sponsoring the research. Much less attention has been given to how polling as a component of the political process impacts representation, or how it is utilized by the press within the political context, particularly as it relates to the attention given to reporting on minority interests. Data comes from the Los Angeles Times poll, which dates back to 1977, as well as content analysis of news coverage of ballot initiative and candidate campaigns associated with these polls. It will examine trends in the sampling of minorities, including sampling methodologies, and explore correlations between subgroup sample sizes, majority-minority attitudinal polarization, and news coverage of the opinions of racial and ethnic minorities. This study will also include discussion of the methodological approach employed in the 2016 LA Times Day Break Poll, which made headlines for its unconventional weighting scheme and reputation as one of the few polls that had President-elect Donald Trump in the lead over the course of the campaign.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Innovations in Cognitive Interviewing: Recruitment Strategies and Interviewing Techniques

Concurrent versus Retrospective Think-Aloud Method in 4th Grade Children

Mila Sugovic *EurekaFacts, LLC*
Ismail Nooraddini *EurekaFacts, LLC*
Cecilia Teal *EurekaFacts, LLC*
Bohdana Sherehiy *EurekaFacts, LLC*

During the cognitive testing of survey items a think aloud technique is used to elicit information that allows making inferences about the cognitive processes underlying survey item response. Think aloud requires respondents to verbalize their thoughts as they answer survey questions (concurrent think aloud (CTA)) or after they answer the survey question (retrospective think aloud (RTA)). Although the literature contains guidelines when the CTA versus RTA technique should be applied, there is little to no research investigating how these two approaches affect quality of verbal reports and, thus, survey pretests. Previous research has shown that think aloud affects problem solving quality (Flaherty 1975) and may alter the respondent's understanding of the question (Trenor et al., 2011). Additionally, past research evaluated think aloud techniques during usability testing, but little research has involved survey pretesting (Haak et al., 2003; Peute et al., 2015), in particular with children. This study examines the application of RTA and CTA techniques during testing of a computer-based assessment in grade 4 students (age 9 - 11). We compared the effectiveness of the think aloud approaches (CTA versus RTA) for 4th graders by evaluating the number of reported problems across items, verbal report length, the quality and relevance of the student verbal protocols. We found that the quality and relevance of the verbal reports provided by students varies between the CTA and RTA method. Children who received the RTA method provided shorter and less relevant verbal reports to the question topic, compared to those students who received the CTA method. Therefore, the effect of CTA vs RTA methods on quality of verbal reports has important implications for survey item and usability testing.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Questionnaire Design in the 3MC Context

Age: Cross-National and Cross-Cultural Challenges

Paul Harwood *Twitter, Inc.*

Wilson Chan *Twitter Inc*

This paper addresses the cross-national and cross-cultural challenges faced in constructing response sets for age questions. The public opinion research literature is comprehensive that question wording matters, and in our study we focus our attention to investigate the effects of response set wording. With a greying global population, understanding the effects of age question response set wording (e.g. pro social response etc.) is critical to best understand the diversity of our user base, and the public generally. Utilizing our annual user survey, which is conducted in 13 markets in nine languages, we establish multiple wording conditions to determine what effect cohort versus exact age measurements have cross-nationally and cross-culturally. To determine the effect we use self-reported and survey para-data.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Questionnaire Design in the 3MC Context

Culturally-Related Response Styles for Attitude Questions: A Comparative Analysis of Chinese and American Respondents

Mengyang Wang *University of Nebraska--Lincoln*

Larry Williams *University of Nebraska--Lincoln*

Shanshan Deng *University of Nebraska--Lincoln*

Ana Lucia Cordova-Cazar *University of Nebraska-Lincoln*

Lok Wa Yuen *University of Nebraska--Lincoln*

When conducting cross-cultural surveys, researchers face unique challenges in confronting two different types of variance that may arise: true variance and culturally-related variance. In drawing valid conclusions from survey, researchers must understand the influence these types of variance may have on the response. Culturally-related responses may be influenced by whether an individual is part of a more individualistic culture, such as the United States, or a more collectivistic culture, such as China. This research hypothesizes that when compared to American respondents, Chinese respondents will be more likely to provide responses that are more conforming, more socially approved, and less extreme regardless of the survey content, as a manifestation of collectivistic norm. In this research, a comparative analysis of responses from Chinese and American respondents is performed using data from Gallup World Poll. Latent class analysis is conducted to investigate the differences in the response patterns of Chinese and American respondents to determine whether any variances can be explained by the underlying cultural differences. Specifically, this research seeks to examine six different stylistic responses and to see whether the tendency of Chinese respondents is to acquiesce, to express neutral or mild rather than polarized attitudes, or to endorse the non-opinion response independent of the instrument contents, significantly differs from that of American respondents. Preliminary results indicate a complex pattern where American and Chinese respondents exhibited some unique response styles respectively. American respondents are two times more likely than Chinese respondents to opt for extreme responses across a range of topics. Moreover, Chinese respondents appear, on average, six times more likely than American respondents to select non-opinion response. These findings raise a number of issues which are important to address in order to assess the problem of measurement artifacts unrelated to the constructs of interest before valid cross-cultural comparisons can be made.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Fit Your Purpose: Frameworks and Examples of Alternatives to Probability Sampling

Indirect sampling for RDS designs

Guillaume Filteau *University of North Carolina at Chapel Hill*

Robert Agans *The Univ of North Carolina at Chapel Hill*

Donglin Zeng *University of North Carolina at Chapel Hill*

Hard-to-reach populations, such as drug injectors and sex workers, are characterized by the difficulty or impossibility in sampling them from a known sampling frame. Respondent-driven sampling (RDS) is a methodology that was developed to study such populations by exploiting chain-referrals over their social networks. Unfortunately, because the method depends on an initial sample recruited through convenience sampling, RDS produces biased estimates. To eliminate this bias, we suggest a combination of the RDS methodology with indirect sampling, in which the hard-to-reach population is sampled through an auxiliary referral population with a known sampling frame. Since the initial sample now comes from a known sampling frame, the new RDS design produces exact probability weights. We propose new estimators for the method, which are design-unbiased or nearly design-unbiased. The properties of the estimator are explored in a simulation study, and compared to those of existing estimators. We also apply the method to recruit a lesbian, gay and bisexual (LGB) sample, and compare the results of the new estimators to those of the CDC and GDTS.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Innovations in Cognitive Interviewing: Recruitment Strategies and Interviewing Techniques

Methodological Considerations in the Use of Web Probing for Questionnaire Evaluation

Stephanie L Fowler *National Cancer Institute*

Gordon Willis *National Cancer Inst., NIH*

Richard Moser *National Cancer Institute*

Dana Wolff-Hughes *National Institutes of Health*

Paul Scanlon *Centers for Disease Control and Prevention*

Reanne Townsend *Westat*

Terisa Davis *Westat*

Over the past several decades, methods for questionnaire development, evaluation, and testing have largely involved the conduct of cognitive interviewing. However, because of the time and resources required to conduct traditional lab-based cognitive interviews, several major shifts in this approach have led to internet-based systems that facilitate access to a wide range and large number of respondents, quickly and cheaply. These challenges and opportunities have led survey researchers to consider fundamentally new approaches to cognitive interviewing, and in particular, the development of web probing as a methodology in our toolbox. Therefore, the present investigation examined two critical issues related to web probing. First, to examine the utility of web probes relative to standard cognitive interviews, we tested if these differing administration modes produced similar information. Responses to the cognitive interviews and web probes were coded between two raters for usefulness of responses and themes that emerged. The probes were identical between administration modes and were targeted at items from the National Health Interview Survey (NHIS). The second issue relates to placement of probes, and we examined if length and quality of responses varies by placement (Willis, 2005). We conducted two web-based experiments, using Amazon's Mechanical Turk, and assigned respondents to either a condition where the probes immediately followed the target item (embedded) or a condition where the probes were placed at the end (retrospective). We compared placement differences on 4 probes (NHIS; experiment 1) and 5 probes (Health Information National Trends Survey; experiment 2), respectively, by examining: the number of words entered in response to each probe, percentage of non-response, usefulness of responses (interrater coding), and themes that emerged (thematic coding). Finally, we synthesized the results of these studies to produce a set of integrated conclusions concerning the practice of web probing.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

The Election, Polls and the Media

What the Public Learned about Donald Trump and Hillary Clinton during the 2016 Campaign

Michael W Traugott *University of Michigan*

Frank Newport *Gallup, Inc.*

From early July through November 8, Gallup asked a probability sample of 500 Americans with telephones each night whether they had heard, read, or seen anything about each of the candidates “in the last couple of days.” If a respondent answered “yes,” they were asked what they had heard, and the response was recorded verbatim by the interviewer. Through Election Day, this information was obtained from almost 59,000 respondents. The recall of having read, seen or heard something varied by candidate and over time, and those variations form an important area of analysis in and of themselves. On average, significant majorities responded affirmatively each night. In this paper, we will trace both the content and sentiment of Americans’ responses, tracking how they changed over the course of the campaign in conjunction with critical events like the conventions, debates, Clinton’s health issues, the public release of Trump’s Access Hollywood video, and FBI Director Comey’s two letters to Congress late in the campaign. Preliminary analyses show that rarely during the campaign could Clinton escape recollections about her emails, whether from news coverage of her private server, congressional testimony, Wikileaks disclosures of John Podesta’s hacked emails, or the potential discovery of copies of emails on Anthony Winer’s laptop. For Trump, on the other hand, Americans’ recollections of information about him covered more topics and generally were neutral or at least more favorable than the recollections of Clinton. One exception for Trump was the increase in recall of having heard about the video and his treatment of women. The paper will explore the link between these recollections and the relative favorability of the two candidates, along with partisan and demographic differences in recall.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Questionnaire Design in the 3MC Context

The Legitimation of Inequality, Equity Theory, and Economic Transparency: A Theory and Two Empirical Tests Using a New Survey-Based Measurement Scale

Jonathan Kelley *International Survey Center and University of Nevada, Reno*

M.D.R. Evans *University of Nevada*

Sarah M.C. Kelley *University of California, Berkeley*

The amount of income inequality thought legitimate, a matter fiercely debated, differs greatly from nation to nation. In these debates the predominant legitimating principle is equity theory, the claim that justice lies in rewarding people in proportion to their productivity. This principle is natural in simple economies of small, independent producers where productivity is transparent, like most of the world before the industrial revolution. But in a modern economy of large, bureaucratic organizations with a complex division of labor, or in more complex or less monetized settings within any economy (government, education, religion, the military), individual contributions are often unclear. That gives less justification for either the high pay of productive workers, or the low pay of unproductive workers. Thus equity theory legitimates a relatively egalitarian distribution of income in modern economies and in economic sectors where productivity is not transparent. This theoretical prediction is tested empirically using a new multiple item scale measuring attitudes to productivity, several long established scales measuring attitudes to various aspects of inequality, together with detailed background and socioeconomic measures. Data are from the International Social Science Survey, Round 20, USA 2016 (a large internet based sample of the adult US population). A second test is based on the ISSS Round 18, Australia 2004 (a large representative national sample conducted by mail based on a simple random sample of the compulsory electoral register, with a completion rate over 60%).

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Fit Your Purpose: Frameworks and Examples of Alternatives to Probability Sampling

Why the Total Survey Error framework is wrong for nonprobability surveys

Andrew W. Mercer *Pew Research Center*

The standard Total Survey Error (TSE) approach to what Groves (1989) called “errors of nonobservation” classifies error as attributable to coverage, sampling, or nonresponse according to the stage in the process in which it is introduced. Classifying error in terms of process makes sense for probability-based surveys because they rely on randomization to generalize from sample to population, and the degree to which randomization is successful is a feature of the survey process. The focus on coverage, sampling, and nonresponse is helpful because it directs focus to those parts of the process where deviations from random selection can occur. However, for nonprobability surveys, the traditional TSE error typology falls short. Because we have abandoned random selection, we cannot base our judgements about the reliability of survey estimates on any intrinsic properties of the survey process. Inference in nonprobability samples instead depends on models of the relationship between sampled and nonsampled units in the target population. In this paper, we illustrate how and why the standard TSE approach to errors of nonobservation are inadequate for nonprobability samples. We briefly describe an alternative framework based on principles from the field of causal inference proposed by Mercer, Kreuter, Keeter and Stuart (forthcoming in POQ) that describes selection error in terms of how well a given sample and model jointly replicate the structure of the target population with respect to outcome and adjustment variables. We demonstrate how the features that describe selection error in structural terms – exchangeability, positivity, and composition – overcome the shortcomings of the TSE approach, and review examples of empirical work conducted under this structural framework.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Making American Great: Domestic Policy

Attitudes and misperceptions toward welfare recipients: an updated analysis

Charles Dahan *Stanford University*

Charles Dahan *Stanford University*

Casey C Ste Claire *University of California, Berkeley*

Class inequality in the United States has become especially salient in the wake of the 2016 Democratic primary and general election, but the lowest-income Americans—welfare recipients—were hardly the focus of Bernie Sanders, Hillary Clinton, or Donald Trump. Past research into welfare recipients demonstrates that welfare recipients are portrayed as more working-age and black by the media than they actually are; however, much of this research is difficult to place in the context of a 2016 United States. This paper argues that re-examining research on welfare is compelling, due to many developments in American society that have taken place since the 1990s, when a majority of this research was published. These developments include, but are not limited to, the election of the first African-American president and the emergence of the 24/7 news networks. We intend to employ conjoint analysis as one of our experiments, which allows us to work with respondents who will assess hypothetical preferences as to who receives welfare across a docket of selected attitudes. Multiple hypotheses regarding public perceptions of welfare can simultaneously be tested using this conjoint design, including age and race. Our intended results are threefold: to indicate that work in fixing public opinion on welfare is desperately needed; to shed new light on specific welfare-related attitudes; and to further demonstrate the value and breadth of conjoint analysis in political science.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

Fit Your Purpose: Frameworks and Examples of Alternatives to Probability Sampling

A framework for large scale non-probability polling, with examples from the 2016 presidential election

David Rothschild *Microsoft Research*
tobi kontizer *stanford*

response rates, non-probability polling in combination with high-end post-sampling analytics is becoming more important for gauging public opinion, given advantages in cost and quick turn-around time. However, a rigorous discussion of a general framework for non-probability polls is overdue, causing some concern about the scientific nature of these polls. To this end, we develop the first generalized framework for understanding the sources of error in non-probability polling, akin to the “Total Survey Error” framework. We use this framework to conduct a post-mortem of two non-probability surveys that we ran during the 2016 presidential campaign, one conducted by Pollfish via smartphones and one conducted on MSN.com with a total of over 400,000 responses between January 1 and Election Day. We develop strategies to deal with the non-probability-polling equivalent of coverage error, sampling error and non-response, and develop ways to address potential error sources unique to non-probability polling. In particular, we develop a dynamic post-stratification model allowing us to parse out true swings from swings in sample-composition. Our work shows the promise of large scale non-probability polling, and the considerations that must be addressed to make it viable.

May 20th, 2017
8:00 AM - 9:30 AM
Concurrent Session F

We Miss You. A Little Too Much, A Little Too Often: Nonresponse Prevention, Evaluation and Adjustment

Non-Response Adjustments for SurveyMonkey Election Tracking

Jack Chen *SurveyMonkey*

Sarah Cho *SurveyMonkey*

For this study, we examine respondents to SurveyMonkey's Election Tracking poll, which had more than 10,000 respondents each week from January to November 2016. Respondents were selected from a random sample of respondents to user-generated SurveyMonkey surveys with the opportunity to participate in our election poll. Unlike most surveys, in which nothing is known about the characteristics of non-respondents, the fact that all respondents and non-respondents have already completed a SurveyMonkey survey means that we have some information about both groups. From respondents, we collect demographic information along with information on political affiliation, party identification, and responses to our questions on the 2016 election. From both non-respondents and respondents, we have information on which surveys respondents completed before opting into our Election Tracking poll. In this analysis, we compare the population of registered voters who completed the Election Tracking Poll to the sample from which they were drawn. We propose several adjustments for non-response, based on demographics, political factors, and paradata (device type, completion times, prior survey type).

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

New Insights on Interviewer Effects in Surveys (Panel)

New Insights on Interviewer Effects in Surveys

Simon Kühne *Socio-Economic Panel (SOEP, DIW Berlin)*

Brady West

Brady West *University of Michigan*

Zeina Mneimneh *University of Michigan*

Julie De Jong *University of Michigan*

Mansoor Moaddel *University of Maryland*

Sharan Sharma *University of Michigan*

Michael Elliott *University of Michigan*

Dana Garbarski *Loyola University Chicago*

Nora Cate Schaeffer *University of Wisconsin-Madison*

Jennifer Dykema *University of Wisconsin-Madison*

Despite the growing importance of digital and online technologies, interviewers are still playing a major role in survey data collection. While interviewers can clearly contribute to participation rates and data quality, they can also negatively influence response behavior through their mere presence, their characteristics, as well as their interviewing behavior. These interviewer effects are present in almost all types of surveys and questions and generally threaten survey data quality by introducing variance and bias into estimates. Although survey researchers have been studying interviewer effects for nearly a century, the exact underlying mechanisms causing these effects still remain unclear, and many questions have not yet been sufficiently answered: What is the role of interviewers vs. respondents in the occurrence of interviewer effects? How do interviewer characteristics beyond socio-demographics, such as personality and attitudes, affect responses? How can we identify and reduce interviewer effects during actual fieldwork? This session will address ongoing issues and knowledge gaps in research on interviewer effects. New insights on the determinants and the nature of interviewer-related error will be discussed. The five contributions in this panel will cover a variety of topics, such as the impacts of interviewer characteristics and behaviors on total survey error, interviewer observations, and interviewer quality control. Collectively, this panel aims to set a new agenda for future research on interviewer effects.

Where do we go from here? Future directions for research on interviewer effects based on a comprehensive research synthesis

Brady West

For nearly 100 years, survey researchers have investigated explanations for variable interviewer effects on survey data using a variety of study designs and analytic approaches. When using the total survey error framework to synthesize this vast literature, we can learn a lot about what we know (and what we don't know) regarding the variable effects that interviewers can have on coverage, nonresponse, measurement, and processing errors. This presentation will highlight important gaps in knowledge and necessary directions for future research on interviewer effects based on a comprehensive review of this literature. Briefly, needs exist for more structured mediation analyses, clever approaches to designing studies that examine interviewer variance in various types of bias (as opposed to coverage,

nonresponse, or processing rates), and more work on the effects of interviewers on multivariate statistics of interest.

Toward a Better Understanding of Interviewer Effects in a Nationally Representative Survey in Tunisia

Zeina Mneimneh, Julie de Jong, Mansoor Moaddel

Research has shown that interviewers can have important effects on respondent answers. Potential bias introduced by interviewer gender and religious wardrobe on related survey items is of particular concern in the gender-segregated and religious context of the Middle East and North Africa. For example, studies in the region have found that male respondents reported more egalitarian views to female interviewers (Benstead, 2013) and that interviewers wearing Islamic (rather than secular) symbols and Islamic hijab (vs. no hijab) received increased reporting of religious attitudes either directly or through an interaction with respondents characteristics (Blaydes & Gillum, 2013; Benstead, 2014; Koker, 2009; Mneimneh et al., 2015). Moreover, we have recently shown that an interviewer's own religious attitudes affected respondent's reported religious attitudes independent of interviewer religious wardrobe. The effect of an interviewer's attitudes was as large as, and sometimes larger than, the effect of the interviewer's religious wardrobe (Mneimneh et al., 2015). The literature, however, is lacking on an explanation of the mechanism of these effects. Are interviewers mirroring the attitudes of the respondents they are interviewing or are they projecting their own attitudes on the respondents? Are the effects transmitted through potential side conversations about religious topics between the respondent and the interviewer? Using recently available panel survey data from a second wave of data collected in Tunisia in 2015, this paper investigates these research questions by looking at interviewer's attitudinal measures collected before the field work and contrasting their effects with interviewer measures collected after the field work. Moreover, observational measures on side conversations related to religious and political topics were collected, allowing for investigation of the potential mediating or moderating effects on the relationship between interviewer's and respondent's attitudes.

Interpersonal Inferences and Interviewer Effects in Face-to-Face Surveys

Simon Kühne

It is well known that the presence of an interviewer can affect responses and thereby introduce variance and bias into survey estimates. For instance, some respondents tend to adjust their true answers towards social norms or specific characteristics of the interviewer in order to appear in a good light. When investigating these interviewer effects, survey research mainly focused on interviewer socio-demographics and only a few studies have examined effects of not directly observable characteristics such as interviewer personality, attitudes and beliefs. Moreover, survey research lacks of insights on how interviewers' and respondents' interpersonal perceptions of each other affect respondent answers to related questions. For this project, self-reports of 1,184 respondents and 114 interviewers as well as their mutual perceptions of each other were collected in the context of the Socio-Economic Panel Study Innovation Sample (SOEP-IS) in 2015/2016. Data collection covered attitudes and opinions towards a variety of political and social issues. This presentation includes results on the impact of interviewer opinions on respondent answers, the nature and accuracy of interpersonal inferences, as well as their impact on respondents' self-reports.

Predicting the interviewers behind interviewer effects: Exploring the utility of computer-generated paradata to set up an active interviewer monitoring system

Sharan Sharma, Michael Elliott

Literature on interviewer effects is largely devoted to their estimation and understanding their causes and associated factors. However, consideration of interviewer effects in active quality control does not seem widespread, despite knowing that it can greatly reduce precision of survey estimates. Similarly, computer-generated paradata (CGP) are underutilized in regular survey practice despite their promise. We explore if CGP can be used in a more principled fashion in the context of a formal interviewer monitoring system. The final goal is to find interviewers who may be contributing to measurement error for specific items so that intervention can be targeted and tailored during active fieldwork. We focus on repeated cross-section surveys in this talk, using data from a face-to-face survey in India and a telephone survey in the U.S to address our questions.

Examining the Validity of Interviewers' Ratings of Respondents' Health

Dana Garbarski, Nora Cate Schaeffer, Jennifer Dykema

Self-rated health – e.g., “would you say your health in general is excellent, very good, good, fair, or poor?” – is widely used to study the health of respondents in surveys because it is a good predictor of subsequent morbidity and mortality. However, because this question allows respondents to consider their health across the multitude of domains salient to them, systematic differences in reporting across population subgroups often make it difficult to compare self-rated health across subgroups. Thus, researchers seek to capitalize on the strengths of the self-rated health measure while mitigating this weakness of systematic reporting differences. Global health ratings about a respondent made by someone else may provide additional and complementary information about the respondent's health status. Recent research has found that this may be particularly true when ratings are made by non-medical personnel such as an interviewer. Incorporating interviewers' assessment of the respondents' health might be a cost-effective measure of health and is consistent with the increasing use of interviewer observations in survey research. However, a better understanding of what factors contribute to interviewers' assessments of respondents' health is needed. This study seeks to establish the criterion validity of interviewers' ratings of respondents' general health status (IRH), examining associations between IRH and health- and interviewer-relevant measures of interest. We analyze data from two studies: Wave 8 of the UK Innovation Panel Study and the 2016 wave of the US General Social Survey. Importantly, while IRH has been exclusively measured at the end of the survey interview in previous research, we experimentally manipulate when interviewers rate respondents' health—at the beginning or the end of the interview—to examine whether the associations between IRH and criteria vary depending on IRH's placement. The results have implications for when researchers should ask interviewers to make assessments about respondents during

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Poll and Poll Aggregation Challenges During the 2016 Election Cycle (Panel)

Organized by DC-AAPOR

Poll and Poll Aggregation Challenges During the 2016 Election Cycle

Gina Walejko *U.S. Census Bureau*

David Rothschild

David Rothschild *Microsoft & PredictWise*

Natalie Jackson *Huffington Post*

Tyler Sinclair *Morning Consult*

Harry Enten *FiveThirtyEight*

Courtney Kennedy *Pew Research Center*

Gina Walejko *U.S. Census Bureau*

Throughout the 2016 election cycle, statistical methods associated with polls and poll aggregators evolved in an attempt to predict more accurately the outcomes of contested races. This panel brings together the people who ran polls and poll aggregation models to discuss polling challenges in the context of 2016 polling wins and misses. The first two talks discuss poll aggregation methods and models during 2016. Over the last few election cycles, poll aggregation using various data and techniques has flourished. Some aggregators use state-polls only, while others add in national polling, fundamental polling, and market data. Aggregation techniques range from simple averaging with a regression model to complex Bayesian models combined with poll-level effects. Presenters discuss how poll aggregation models differed, the challenges associated with different techniques, and what poll aggregators should do differently in future election cycles. The second two talks discuss the “shy Trump effect.” Using different datasets, the presenters explore the hypothesis that Trump supporters were reluctant to reveal their support in polls. Presenters explore this hypothesis using aggregated data as well as an embedded mode experiment from a series of polls. The final talk discusses whether pollsters should weight to party affiliation. The presenter describes the results of a study examining several weighting adjustments designed to reduce differential partisan nonresponse tested on data from the 2012 and 2016 presidential elections.

Poll Aggregation: Looking Forward

David Rothschild

The 2016 election was a great test of the aggregation of polling. Topline polling performed fairly well on a state-by-state outcome, pointing in the winning direction for 46 or 47 of 51 Electoral College elections, with an average error that was slightly larger than normal. However, error was highly correlated and large for a band of Midwestern states, leading to a surprising overall election outcome. This talk argues that – as long as poll aggregation is confined to topline-level polling data – it is impossible to accurately reflect, let alone correct, for underlying correlated bias. Looking forward, it questions how poll aggregation will adapt without having access to voter file and individual-level polling data.

How Much Should Individual Polls Matter In Aggregation and Forecasting?

Natalie Jackson

HuffPost Pollster uses model-based approaches to aggregating polls, and for general election polls we use a Bayesian time series model. This was the baseline for our forecast model. This talk will discuss the advantages and disadvantages of using this type of model, as well as the parameters that can be adjusted to make such a model more or less responsive to individual polls. The talk will also address the decisions we made about state correlations, national polls' influence, and whether our criteria for including polls affected the model. The talk will conclude by describing plans for the future.

Dismissing the “Shy Trump” Effect

Harry Enten

The 2016 election saw Donald Trump winning, despite polls indicating he would lose. Why did that happen? While polls indicate national surveys did about as well as they have over the past few decades, state polls had a wider margin of error than usual. This talk explores whether a “shy Trump” effect was to blame. It discusses a study of the states with the largest errors and reveals that this effect was probably not to blame. Well-educated states, where Trump voters would seemingly be most reticent to reveal support, had some of the lowest errors. Red states, where Trump supporters would be the most likely to feel safe in revealing support for Trump, suffered the largest errors.

Survey Mode Effects During the 2016 Election Cycle

Tyler Sinclair

Morning Consult conducted a number of studies throughout the 2016 election that helped to explain Donald Trump’s rise in the primaries and his relatively stable levels of support in national polling during the general election campaign. This presentation will highlight results from three such studies: 1) a December 2015 survey mode effects experiment during the Republican primary where we found a significant difference in support for Donald Trump in online versus phone surveys, 2) an October 2016 re-run of the mode effects experiment where we found a small overall mode effect, but some pockets of “Shy Trump” voters, and 3) a comprehensive exit poll including more than 25,000 voters who were asked more than 100 questions on key policy and political topics.

Adjustments for Differential Partisan Nonresponse in Public Opinion Surveys

Courtney Kennedy

One of the more contentious questions facing public opinion researchers is whether they should be aligning their survey samples to estimated population parameters for political party affiliation – in other words – whether to weight to Party ID. The potential upside of weighting to party is the reduction of bias stemming from differential partisan nonresponse. Among the downsides of weighting to party, however, are the fact that party is an attitude rather than a demographic and the lack of authoritative benchmarks. Both of those features suggest that weighting to party may entail a fair amount of measurement error, adding noise to weighted estimates. Currently about half of election pollsters adjust for party affiliation in their weighting, and this half of the field is predominantly (though not exclusively) those fielding their surveys online. Techniques for adjusting for party vary across researchers and include using party as a raking dimension, multi-level regression and post-stratification, as well as using party in a propensity model. In this presentation, we share the results of a study that examined several weighting adjustments that were designed to reduce differential partisan nonresponse. One adjustment tested raking to an enhanced set of demographic variables correlated with party, including cross-classification of state (or grouped states) with race, age, sex, and education (separately). Another adjustment tested raking to an ecological partisanship measure, based on the past presidential vote

margin in the respondent's county. These adjustments were compared to a weighting protocol that included party affiliation in the raking in a straightforward fashion. We tested the adjustments on data from the 2012 and 2016 presidential elections.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Reaching the Hard to Reach: Insights and Solutions

Internet and Technology Usage by Households Applying for Free and Reduced-Price School Meals

Jessica E Graber *US Census Bureau*
Alfred Tuttle *U.S. Census Bureau*
Lucia Lykke *U.S. Census Bureau*
Kathleen Kephart *U.S. Census Bureau*

This paper describes research aimed at improving measurement of household eligibility for participation in school meals programs. The Food and Nutrition Service of the United States Department of Agriculture oversees the largest federal program for school-age children, which in 2012 provided more than 31 million lunches and 10 million breakfasts each school day. Eligibility for the program can be determined in multiple ways including covering entire communities based on local poverty levels, household participation in other public assistance programs, and household-level applications for one or more child. Household applicants are asked to provide detailed household roster information and sources of income for all household members. Applications have traditionally been completed using paper-based forms that are processed locally. The Food and Nutrition Service is now developing an internet application option that could be administered by local agencies. Electronic applications provide an opportunity to increase data quality by reducing missing data and inconsistencies, but the viability of this option for these households is unknown. As part of an effort to reduce household application errors, we conducted detailed, in-person interviews with more than 150 adults who applied for free or reduced-price school meals for children in their households during the 2016-2017 school year. Respondents were asked about internet access in their household and, if available, their specific method of internet access (mobile v. desktop/laptop). We also asked about cellular or smartphone use. We collected information on respondents' online behavior and preferences, including conducting financial, business and other official transactions. This research provides important insight on attitudes towards internet use among lower SES households. For the Food and Nutrition Service, our findings will also inform planned usability testing of the internet application to ensure that the design and function of the tool meets the needs of the end user.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Over Rated or Under Weighted? Methods for Improving Inferences from Online Nonprobability Samples

www.Are_You_In_My_Sample.com: A Deeper Dive on the Digital Divide

David Dutwin *SSRS*

Trent Buskirk *U. Mass Boston*

Sample survey costs matter, perhaps more now than ever, and many practitioners see the move to online surveys as one way to reduce costs. In fact, some cite the attraction to probability panels is related in part to lower costs and similarly for ABS samples that push respondents to the web. However, in an attempt to view the choice of online surveys or not within the total survey error context, one still needs to balance cost implications with possible errors associated with coverage. Currently, according to many publically available sources, the percentage of U.S. households with internet access lies between 85 and 88 percent depending on how internet access is measured. If those who have internet access differ from those who don't, then surveys relying only on respondents with internet access may suffer from coverage biases. Such biases might be attenuated if an explicit correction for internet access could be made. In this paper we explore differences between respondents with and without internet access and then quantify potential biases related to internet coverage using data obtained from a dual frame RDD probability sample. We first analyze and document 16 major publically available surveys including the ANES, GSS, and others on internet access against demographic, technographic, behavioral and political and privacy variables to note which, controlling for demographics, remain as significant drivers of Internet use. We then field a dual-frame telephone survey with those variables that were most significant from the 16 survey analyses, and then construct an internet propensity model via logistic regression based on this core set of predictors. We apply the propensity adjustments to respondents with internet access and compare resulting estimates to the full dual frame sample on a collection of outcomes to explore the adjustment model's utility.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Driving Them to the Web: Strategies, Techniques and Innovations

Testing the Impact of Mail Materials on Web Participation in the National Immunization Survey

Benjamin Skalland *NORC at the University of Chicago*

Jacquelyn George *NORC at the University of Chicago*

Vincent Welch *NORC at the University of Chicago*

Holly Hill *Centers for Disease Control and Prevention*

Laurie Elam-Evans *Centers for Disease Control and Prevention*

Cindi Knighton *Centers for Disease Control and Prevention*

Chalanda Smith *Centers for Disease Control and Prevention*

Declining landline coverage rates, poor mobile phone response, and geographic uncertainty related to mobile phones have made address-based sampling an attractive alternative to random digit dialing (RDD) for survey data collection. A web-first design is one method of address-based sampling that shows a great deal of promise. In this design, respondents are invited to complete the survey online via a mailed invitation. The success of this design hinges on the ability to entice respondents to log on to the web site. As part of the National Immunization Survey (NIS) and the National Immunization Survey – Teen (NIS-Teen), surveys used to monitor vaccination coverage rates among children (19-35 months) and teens (13-17 years) in the United States, respectively, an investigation was conducted to determine the mail procedures and materials that would lead to the highest web login rate. The analysis included 169,000 sampled addresses and examined factors such as type of postage used (first class vs. non-profit), the use of an advance postcard, the text and image on the advance postcard, the inclusion of a web instructions insert, and the use of a “last ditch” postcard. The outcome was the web login rate. We will present the study design, study results, recommendations for future studies using mail to recruit web survey respondents, as well as potential next steps for the NIS.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Adaptive/Responsive Design: Predictions and Evaluations

Transitioning an In-Person Longitudinal Survey to a Mixed-Mode, Two-Phase Survey Design:

Preliminary Results

Paul Biemer *RTI International and University of North Carolina at Chapel Hill*

Kathleen MullanHarris *University of North Carolina*

Brian Burke *RTI International*

Kathleen Considine *RTI International*

Carolyn Halpern *University of North Carolina*

Chirayath Suchindran *University of North Carolina*

The National Longitudinal Study of Adolescent to Adult Health (Add Health) follows a nationally representative sample of about 19,800 adolescents selected from grades 7-12 in the 1994-95 school year. The study is conducting its fifth wave of interviews in 2016 with Wave IV having concluded in 2008. All prior waves were conducted by in-person administration using a combination of computer assisted personal interviewing (CAPI) and computer assisted self-interviewing (CASI). To reduce costs, Wave V is transitioning to a mixed-mode, two-phase survey design. Phase 1 is a mail survey conducted by computer assisted web interviewing (CAWI) and paper and pencil interviewing (PAPI). Phase 2 selects a subsample of nonrespondents for CAPI/CASI field interviewing and allocates the remaining nonrespondents to computer assisted telephone interviewing (CATI) followup using an abbreviated questionnaire. The Wave V sample is being interviewed over a three-year period with a random subsample being interviewed each year. In the first year of interviewing, an experiment is being conducted to test the new mixed-mode, two-phase design, several different incentive levels and a modular questionnaire design. The experiment is designed to address a number of key research questions that will inform the design in Years 2 and 3, including: a) What response rates can be achieved for each phase of the two-phase design? b) Which questionnaire administration produces the higher response rates: a single, 60-minute questionnaire or a two-part modular questionnaire, each of which is 30 minutes in length that are administered separately and sequentially? and c) What level of incentive produces the highest response rate; do response rates differ by questionnaire version? The paper describes the Add Health Wave V survey design and provides some preliminary results from the Year 1 experiments. Implications for general in-person surveys transitioning to self-administered modes are provided.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Measuring and Evaluating Nonresponse

Non-response bias in a dual frame phone survey, are the later respondents different are from early respondents, BRFSS 2015

Pranesh P Chowdhury *Centers for Disease Control and Prevention*

Carol Pierannunzi *Centers for Disease Control and Prevention*

Machell Town *CDC*

William Garvin *CDC*

Declining response rates may increase non-response bias in telephone surveys. Increasing the number of attempts to reach respondents has been shown to increase the response rates and potentially lower non-response bias. If respondents who complete a survey after a small number of attempts are different than those who require more effort, bias may be introduced by restricting the number of attempts. To assess the non-response bias in a dual frame (cell and landline) telephone survey, we used data from 2015 Behavioral Risk Factor Surveillance System (BRFSS) to compare the early responders (completed the interview with in first, second, or third calling attempts) with later responders (completed the interview with four to six and seven or more calling attempts) on demographic factors, health risk behaviors, and chronic disease or conditions. Currently the minimum number of attempts for the BRFSS is fifteen for the landline and eight for the cell phone samples. About 52% of all adults were early responders; in landline survey 47% and in cell phone survey 55% adults were early responders. Early responders were significantly more likely to be 65 or older, but significantly less likely to be Hispanics, and Black non-Hispanics in both landline and cell phone surveys than those who required more attempts. Respondents who had lower levels of education, did not have health insurance (age 18-64 years), and were current smokers were significantly more likely to be later responders in the cell phone sample. Increasing the numbers of attempts in a cell phone survey might reduce potential bias by including a larger proportion of Hispanics, Black non-Hispanics, and people with lower educational attainment.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Whet Your Appetite - Survey Data Collection Using Smartphone Apps

Enriching an Ongoing Panel Survey With Mobile Phone Measures: The MoDeM study

Mark Trappmann *Institute for Employment Research*

Sebastian Baehr *Institute for Employment Research*

Georg-Christoph Haas *Institute for Employment Research*

Florian Keusch *University of Mannheim*

Frauke Kreuter *University of Maryland*

The PASS panel survey is a major data source for labor market and poverty research in Germany with annual interviews since 2007. In autumn 2016, the supplemental study MoDeM (Mobile Device Measures) has been set up, in which selected respondents were asked to install a study app on their smartphones. MoDeM combines short questionnaires that can be triggered by geographic location with passive data collection on a variety of measures (e.g. geographic location, app use). The triggering of questions allows us to enrich yearly retrospective data with data collected immediately after a certain event (e.g. placement officer visit). Passive data collection allows innovative measures (e.g. for social capital) that complement traditional survey measures. Furthermore the additional smartphone measures create the potential to address new research questions related to the labor market and technology use (digital stress, home office performance). Finally, the study will provide new insights in the day structure and coping behavior of unemployed persons and thus replicate aspects of the classic Marienthal case study with modern means. In this presentation we will give an overview of the study. We will focus on data protection issues, implementation of the fieldwork, selectivity of participation in the study and of participation in short surveys under different incentive schemes. Furthermore, we will assess validity of the new measures compared to traditional survey measures.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Adaptive/Responsive Design: Predictions and Evaluations

Aggressive, Relaxed, or Simply the Default? Adaptive Survey Design Strategies to Reduce Nonresponse Error

Nicole McDermott Tate *RTI International*

Antje Kirchner *RTI International*

Emilia Peytcheva *RTI International*

Jennifer G Cooney *RTI International*

Natasha Janson *RTI International*

Adaptive designs have been used for almost a decade in survey research, providing researchers with methods to balance survey quality and survey cost during data collection (Groves and Heeringa, 2006). This article evaluates the effectiveness of such a design aiming at increasing response rates and reducing the potential for nonresponse bias in the 2016/17 Baccalaureate and Beyond Longitudinal Study (B&B, field test) of college graduates. We use prior wave response behavior to assign cases to different data collection protocols, varying contact frequency, mode, incentive amount, and interview length. More specifically, prior wave nonrespondents were randomly assigned to either a default (used in the prior data collection), or an aggressive data collection protocol (aimed at converting nonrespondents through higher incentives, more frequent contacts, and an abbreviated interview). Prior study early respondents were assigned to a relaxed protocol, and all other respondents to the study default protocol. Data collection consisted of three phases, where different design components (incentive, survey length, or mode) were changed. We investigate response rates, sample representativeness and nonresponse bias using frame data for each group and each phase. The availability of a rich set of auxiliary frame data for all sample members allows us to compare the effectiveness of the different data collection protocols. Preliminary results indicate the response rate is significantly higher in the aggressive ($RR2=34.3\%$) compared to the default ($RR2=19.8\%$, $p<0.001$) data collection protocols for the base-year nonrespondents, and significantly higher in the relaxed ($RR2=71.0\%$) compared to the default ($RR2=63.5\%$, $p<0.01$) data collection protocols for the base-year respondents. Furthermore, while response rates significantly increase with each data collection phase, nonresponse bias assessed using Mahalanobis distance measures decreased significantly in the first data collection phase, but not in later phases. We discuss implications of our findings for data quality and efficiency gains for data collection.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Reaching the Hard to Reach: Insights and Solutions

Decennial Census Knowledge & Participation across Hard-to-Count Sub-Groups

Yazmin A Garcia Trejo *U.S. Census Bureau - Washington, DC*

Gina Walejko *U.S. Census Bureau*

This paper examines the relationship between knowledge about and participation in the decennial census among populations who are hard-to-count (e.g. people with language barriers and/or socioeconomic disadvantages). Past research among hard-to-count populations uncovered a relationship between decennial census knowledge and participation. For example, low awareness scores and low familiarity with the 2010 census were associated with lower likelihood of participation for non-English speakers (Bates and Pan 2010). Other ongoing research focuses on how census knowledge (e.g. benefits, misconceptions and legal knowledge) varies among Hispanic sub-groups and impacts Hispanic census participation (García Trejo 2016). In addition to this, political scientists have found a relationship between factual political knowledge and civic participation (Delli Carpini and Keeter 1996; Pérez 2015). We extend these studies by analyzing data from the Census Barriers, Attitudes and Motivators Survey (CBAMS II) conducted by the Census Bureau in 2011. CBAMS II asked the general population and oversamples drawn from hard-to-count geographies (i.e. rural, high Hispanic density, and high Asian density tracts) questions about census knowledge, attitudes, and participation in the 2010 census. We use factor analyses to uncover knowledge typologies (e.g. benefits and misconceptions about the decennial census) and compare factor scores across hard-to-count sub-groups. We foresee two implications from this research. First, it will expand the understanding of the relationship between census knowledge and participation. Second, it will contribute to the design of messaging strategies aiming to motivate sub-groups of hard-to-count populations to participate in the 2020 census. Findings will also help to inform the design of a third CBAMS-like survey planned to be conducted before the 2020 Census that will inform the use of household-level models targeting addressable media to a diversity of audiences.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Whet Your Appetite - Survey Data Collection Using Smartphone Apps

Using an App to Collect Detailed Expenditure Data in a Probability Household Panel Survey: Response Rates, Response Biases and Measurement Quality

Annette Jackle *Univ. of Essex*

Carli Lessof *University of Southampton*

Jonathan Burton *University of Essex*

Mick Couper *University of Michigan*

In this paper we examine the quality of data on household expenditure obtained using a receipt scanning App in a nationally representative household panel study. We invited 2,432 members of the UK Household Longitudinal Study Innovation Panel to download an App to report all their spending on goods and services for a month. Respondents were asked to each day scan all receipts, report spending for which they did not have receipts, or report that they had not made any purchases that day. The incentive scheme included an experiment with a conditional incentive of £2 versus £6 for downloading the App, followed by a daily conditional incentive of £0.50 for using the App, and a £10 bonus for using the App every day. The App data collection is will be completed early December 2016. We will examine and report on various aspects of data quality. Firstly we will examine participation rates at various stages of the data collection: What proportion of the sample had the required technology (a smartphone or tablet) to participate in the task? What proportion had said they were hypothetically willing to participate in such a task? What proportion actually downloaded the App? How did participation evolve over the month? What impact did the conditional incentive for downloading the App have? Secondly we will examine biases in the types of sample members who participated, considering social-demographic characteristics, financial position and behaviours, mobile device ownership, usage patterns and self-rated skills, and indicators of cooperativeness collected in previous waves of the panel. Third we will examine estimates of monthly individual and household expenditure, comparing the App data with expenditure data from previous waves of the panel, as well as drawing on respondent self-assessments of the proportion of expenditure missed due to non-participant household members, drop-out and under-reporting.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Measuring and Evaluating Nonresponse

Coverage and non-response biases in the National Youth Tobacco Survey

Ronaldo Iachan *ICF International*

Sean Hu *Center for Disease Control and Prevention*

Katherine Flint *ICF International*

Linda J Neff *Center for Disease Control and Prevention*

The National Youth Tobacco Survey (NYTS) is designed to provide data necessary to support the design, implementation, and evaluation of state and national tobacco prevention and control programs. The universe for the study consists of all public and private school students enrolled in regular middle schools and high schools in the United States. A national probability sample is used to select schools and students. This paper examines the sources of potential biases of non-response and under-coverage and assesses their potential magnitude. Under-coverage may result from the exclusion from the frame of alternative schools, special education schools, schools operated by the Department of Defense and Bureau of Indian Affairs, vocational schools that serve only pull-out populations, as well as students enrolled in regular schools unable to complete the questionnaire without special assistance and students in very small regular schools (< 40 students total enrollment). Non-response bias may result from differences in key outcomes between participating and non-participating schools and between participating and non-participating students. We examine response rates at both school levels historically in subgroups defined by school level, school type and other characteristics. These bivariate analyses of non-response inform the multivariate logistic regression models (or propensity models) for the response indicator. The propensity models allow the assessment of the separate effects of each characteristic on response propensity while adjusting for the others. These models, in turn, inform the development of weighting classes for effective non-response weight adjustments designed to minimize the potential biases. We assess the bias reduction gains of weight adjustments by comparing weighted estimates, prior to any post-stratification, to known population totals and percentages. We also assess the variances induced by unequal weighting effects.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Race, Religion, Sex and Gender

Race, Religion, And The Meaning of Evangelicalism

Claire Gecewicz *Pew Research Center*

Claire Gecewicz *Pew Research Center*

Jessica Hamar Martinez *Pew Research Center*

Gregory A Smith *Pew Research Center*

Many surveys categorize Protestants into one of three categories: white Protestants who say "yes" when asked if they consider themselves "born-again or evangelical" Christians are categorized as white evangelical Protestants, white Protestants who say no to this question are categorized as white mainline Protestants, and black Protestants are grouped into a third category regardless of whether they consider themselves born-again or evangelical Christians. When grouped this way, white evangelicals are among the most Republican constituencies in the electorate. White mainline Protestants also tend toward Republicanism, though not to the same degree. And black Protestants are among the most reliably Democratic constituencies in the electorate. In recent years, especially in 2016, this approach has been questioned and criticized for ignoring the racial diversity within Protestantism. After all, observers have noted, not all evangelical Protestants are white. Indeed, roughly two-thirds of black Protestants say "yes" when asked if they consider themselves "born-again or evangelical" Christians. And if polling reports focused their analyses on all evangelicals, not just whites, the political profile of evangelicalism might look very different. This paper uses data from Pew Research Center polls to investigate what difference it makes to subdivide evangelical Protestantism by race. In what ways are African-American evangelical Protestants similar to their white counterparts, and in what ways are they more similar to black Protestants who are not evangelicals or to African Americans more broadly? And what impact would combining African-Americans and other racial and ethnic minorities who self-identify as evangelical Christians with their white counterparts have on our understanding of what it means to be an evangelical – politically, demographically, and religiously? This analysis will help researchers better understand the complicated relationship between race, religion and politics and determine how best to analyze and present their data.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Driving Them to the Web: Strategies, Techniques and Innovations

How much does a promise of a \$5 Gift Card Buy for a Web Survey of College Students? Probably more than you think.

David Cantor *Westat*

One of the most common results in the survey methods literature is that small pre-paid incentives are more effective than promised incentives (e.g., Dillman, et al., 2009). For example, a recent meta-analysis found that a \$2 pre-paid incentive raised the response rate by about as much as a \$30 promised incentive for a telephone interview (Mercer, et al, 2015). However, many surveys are not in a position to provide a pre-paid incentive to everyone in the sample (e.g. web surveys using e-mail addresses). Even RDD telephone surveys are generally unable to do so because of problems with matching addresses to phone numbers. Much of the research on promised incentives are with general population samples using 'traditional' modes of interviewing such as mail, telephone or in-person. There is not as much research on new technologies, such as web surveys, where issues such as dropout rates and item missing data are prominent. This paper presents results of an experiment which tested the use of a promised \$5 gift card on a web survey of college students. Students were told they would receive a \$5 card once they completed the survey. The incentive significantly increased the response rate by 9 percentage points, from 16% to 25%. There was some indication that the incentive had a small, but significant, effect on the types of people who responded to the survey. Perhaps more dramatically, those who were offered the incentive were about 2.5 times more likely to complete the entire survey and were much more likely to complete some of the more sensitive parts of the survey. The results emphasize that 'one size fits all' for incentives is not appropriate. The population and mode of interview should be considered when weighing the costs and benefits of different types of incentives.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Adaptive/Responsive Design: Predictions and Evaluations

Evaluation of a Two-Phase Design for Nonresponse on the California Health Interview Survey

Andy Peytchev *Institute for Social Research, University of Michigan*

Emilia Peytcheva *RTI International, Survey Research Division*

Douglas Currivan *RTI International, Survey Research Division*

Matt Jans *UCLA Center for Health Policy*

Declining response rates for random-digit-dial (RDD) telephone surveys threaten the accuracy of survey estimates and the efficiency of data collection. Responsive and adaptive designs can help minimize survey errors while controlling data collection costs. A key strategy in responsive design protocols is to alter key features of the survey design when phase capacity has been reached for the initial design (Groves and Heeringa, 2006). Monetary incentives is one such feature when the main objective is to reduce nonresponse and the risk of nonresponse bias, as higher incentives have been found to elicit participation from those who are less interested in the survey topic. Informed by prior experimentation in a dual-frame telephone survey setting (Peytchev, Baxter, and Carley-Baxter, 2009), the California Health Interview Survey (CHIS) was redesigned to include two phases of data collection starting in 2015. The redesign included a main data collection phase followed by a more intensive treatment of a subsample of nonrespondents. The second phase used stratified random samples of cell phone and landline nonrespondents drawn with unequal probability of selection, depending on design strata, screener completion, and domains (e.g., ethnicity) determined in the initial phase. This sampling design was adjusted to balance multiple survey objectives. The protocol was altered in the second phase to include significantly larger monetary incentives. This paper evaluates the CHIS 2015-2016 two-phase design in three ways. First, we examine the impact of the second phase on overall response rates and interview yield for each interview type (adult, adolescent, and child). Second, key survey estimates are compared between first phase data and combined data, to assess whether and how the additional phase affected the estimates. Third, we consider the cost effectiveness of the two-phase design by calculating the cost ratio for completing CHIS interviews in the second phase versus continuing the main phase.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Race, Religion, Sex and Gender

Can Respondent Race Alter Perceptions of Events? Biased Processing of Officer-Involved Shootings

Josh Pasek *University of Michigan*

Hakeem Jefferson *University of Michigan*

Fabian Neuner *University of Michigan*

Mounting evidence suggests that individuals selectively expose themselves to political information and process that information in line with partisan identities. Yet it stands to reason that partisanship should not be the only social identity that can alter information seeking and processing. This study examines how biased information reception and processing might account for diverging perceptions of officer-involved shootings among Black and White Americans. Three explanations for this gap are proposed: respondents may select information that bolsters their identities, may engage in motivated reasoning processes to render summary judgments favorable toward ingroup members, or may engage with the facts of each case by updating their views from very different prior beliefs. When Black and White respondents were presented with consistent information about an invented Black-on-White officer-involved shooting, we were able to replicate the sorts of perceptual divides that emerged following a similar incident in Ferguson, MO. This provided evidence that differences in perceptions were attributable to information processing, rather than just information reception. A lack of relations between identity strength, which was primed, and perceptions of the incident also suggested that differences in processing were not a function of rationalization and motivated reasoning. Instead, differing perceptions were attributable to the importance individuals accorded to different pieces of information, which was itself related to respondents' racial biases and their prior beliefs about how police treated individuals of different races. The racial divide therefore appears to stem from preexisting attitudes and beliefs, which change the way new information is processed.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Reaching the Hard to Reach: Insights and Solutions

Correcting for the multiplicity issue in a probability sample of homeless youth

Daniela Golinelli *Mathematica Policy Research*

Joan STucker *RAND Corporation*

William G Shadel *RAND Corporation*

Obtaining probability samples of hard to reach populations such as the homeless youth is challenging for several reasons. One of the primary reasons is that it is difficult to generate a complete sampling frame listing all the members of the population of interest. Therefore, statisticians often resort to alternative sampling methods such as location sampling. Location sampling consists of developing a list of locations where homeless youth, in our study, are known to hang out. However, location sampling introduces other challenges, the primary being something known as the multiplicity issue. The multiplicity issue arises because homeless youth can enter the sample multiple times during the study period. When the population of interest is homeless youth and not homeless youth-visits it is important to devise corrections to the individual sampling weights so that a sample of homeless youth is obtained, instead of a sample of homeless youth-visits. In this paper, we will analyze the effects of these corrections on the estimates of smoking outcomes and background characteristics of homeless youth. In particular, we will compare those homeless youth who visit the sites in our locations frame frequently versus those who don't. We will produce estimates of outcomes and characteristics when no corrections are made for the multiplicity issue and when different types of corrections are made. The results of this analysis will speak to the importance of correcting for the multiplicity issue and how sensitive the results are when these corrections are computed in different ways. The results will also help in developing recommendations for when to use these corrections in other studies that use similar sampling methods.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Measuring and Evaluating Nonresponse

The respondents (and nonrespondents) have spoken! Results of a Rigorous Nonresponse Follow-Up Study Across Multiple Survey Modes

Philip Brenner *University of Massachusetts Boston*

Floyd Fowler *University of Massachusetts Boston*

Trent Buskirk *University of Massachusetts Boston*

Recently surveys have come under fire as a valid tool for social inquiry because of lower levels of response. And while low responses create risk for nonresponse bias, they alone don't imply it. More extensive follow-up studies are needed to understand whether such risks are realized. This research reports on a contemporary example of such a study. Specifically, we examine the potential contribution of nonresponse to total survey error in measures of health care and health and its correlates using a survey that employed multiple modes of recruitment and follow-up. An ABS sample was drawn from five purposively selected Boston neighborhoods/suburbs based on their diversity of demographic and household characteristics. Half of this sample was initially contacted by mail and asked to complete an IVR survey. The remaining sample was divided between two modes. Households on the frame able to be matched to a telephone number were contacted by mail and told to expect a telephone interviewer to call. Those without a telephone number match on the frame were contacted by mail and told to expect an interviewer visit. Samples of nonresponding households from the IVR and telephone modes were then recontacted by mail and told to expect an interviewer visit. Respondents completed a 20-minute interview including a series of questions taken from a number of federally-administered or funded studies. We compare estimates of key health measures (including self-reported health, common medical condition diagnoses, levels of physical activity) from first-round respondents with those from follow-up interviews of nonrespondents, accounting for mode and other design elements.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Race, Religion, Sex and Gender

Does Nonresponse Contribute to Bias in Survey Estimates of Religious Service Attendance?

Philip Brenner *University of Massachusetts Boston*

To what extent are errors in survey estimates of religious service attendance attributable to nonresponse? Since a 1998 symposium in *American Sociological Review*, the debate has continued, but has focused primarily on measurement error, a form of social desirability bias called “overreporting” rather than other potential sources of error. To investigate what role, if any, nonresponse plays in inflating survey estimates of religious behavior, I analyze data from a multi-mode survey designed to allow estimation of nonresponse bias. A sample was drawn from an address-based frame from five purposively selected Boston neighborhoods and suburbs given their diversity of demographic and household characteristics. Half of the sample was initially contacted by mail and asked to complete an IVR survey. The remaining sample was divided between two modes. Households on the frame able to be matched to a telephone number were contacted by mail and told to expect a telephone interviewer to call. A sample of households without a telephone number match on the frame were contacted by mail and told to expect an interviewer visit. Samples of nonresponding households from the IVR and telephone modes were then recontacted by mail and told to expect an interviewer visit. Respondents completed a 20 minute interview including measures of religious behavior, including attendance at religious services. Estimates of religious service attendance from first-round respondents are compared with those from follow-up interviews of nonrespondents, accounting for mode and other design elements. Initial findings suggest some small differences, perhaps primarily due to the success of matching a telephone number to a household, and the demographic characteristics of households with a successful telephone match.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Adaptive/Responsive Design: Predictions and Evaluations

What will work for whom? Identifying subgroups for which response rate interventions will be effective

Michael T Jackson *American Institutes for Research*

Cameron McPhee *American Institutes for Research*

Katrina Steinley *American Institutes for Research*

Leverage salience theory implies that the effectiveness of an intervention aimed at increasing survey response rates is unlikely to be uniform across a sample. Rather, sampled cases are likely to vary in their “sensitivity” to a potential intervention. In principle, an adaptive design that assigns an intervention only to cases predicted to be sensitive to that intervention would allow for a more efficient resource allocation than assigning the intervention to all sampled cases. However, while statistical methods for predicting response propensity are well established, the problem of accurately predicting (in advance of data collection) sensitivity to potential interventions has received limited attention in the survey research literature. This paper will investigate statistical approaches to identifying subgroups that are likely to be particularly sensitive to response rate interventions, using auxiliary data available prior to data collection. Drawing in part on recent advances in predictive subgroup identification from the biostatistics and machine learning literatures, several methods will be compared, including regression-based and recursive partitioning approaches. Data from a randomized incentive experiment incorporated into the 2016 National Household Education Survey—a large-scale ($n = 206,000$) study of U.S. households using an address-based sample and self-administered mailed questionnaires—will be used to demonstrate and evaluate each method. The primary research question is whether any of these methods is able to isolate cohorts between which there is substantial variation in the response rate increase attributable to a \$5 prepaid incentive; and, crucially, whether this information can then be used to accurately predict incentive sensitivity in out-of-sample data. Taking advantage of the large sample size, cross-validation techniques will be used to assess each method’s out-of-sample accuracy. The results will provide insight into whether any of these statistical approaches to predicting incentive sensitivity show promise for use in adaptive survey designs.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Online Polls in the 2016 Election

Hidden Trump Voters? Comfort with Disclosing Candidate Choice to Family, Friends, Acquaintances and Telephone Pollsters in the USC/LA Times “Daybreak” Election Poll

Jill E. Darling *University of Southern California (USC) Center for Economic and Social Research*

Arie Kapteyn *University of Southern California*

During the 2016 presidential campaign, self-administered polls showed higher levels of support for Donald Trump than was seen in interviewer-administered polls, raising the possibility of a social desirability effect. The USC/LA Times “Daybreak” Election Poll team, observing higher support for Trump in our internet poll than was seen in national telephone polls, asked our respondents to rate their level of comfort with disclosing their candidate choice to family, close friends, acquaintances, telephone pollsters, and our own poll (as an upper bound). A sample of 3,167 provided answers that ranged, on average, from a high of 86 for disclosure in our poll, to a low of 54 for disclosure to telephone polls, and analysis revealed significant variation in comfort level between Trump and Clinton voters, and among politically relevant socio-demographic subgroups. Our study provides evidence of variations in comfort levels that could point to social desirability bias effects in interviewer-administered polls. We conclude that it is important to rely on multiple sources of data obtained using different methods and modes during a campaign to help reduce overall error and support more accurate predictions of results on election day. Respondents in this poll were members of the Center for Economic and Social Research’s Understanding America Study (UAS) probability-based internet panel. Daybreak poll members responded weekly from July 4 through November 7th. The comfort with candidate disclosure questions were asked during the week of October 30th. A separate presentation focuses on the methods and outcomes of the Daybreak poll, which used an experimental probabilistic approach successfully employed in forecasting the 2012 election (Gutsche, Kapteyn, Meijer, & Weerman, 2014; Kapteyn, Meijer, & Weerman, 2012) prior to its use this year.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Online Polls in the 2016 Election

Beneath the Toplines: Comparing the demographic and opinion structure of online and live interview RDD samples of 2016 election polls

Charles Franklin *Marquette University Law School*

John DJohnson *Marquette University Law School*

The 2016 election provides an opportunity to compare not just the topline similarity of online with live interview RDD samples, but to also look at the structure of demographic and opinion data collected by each method. We compare SurveyMonkey online samples of Wisconsin taken between July and November with those of the live interview Marquette Law School Poll taken over the same period. SurveyMonkey sampled weekly during this period while Marquette collected 7 statewide samples of 800 to 1400 respondents each. This considerable collection allows comparison over time and with overlapping field periods. Demographic and some opinion items, including vote, are measured identically between the two surveys, allowing estimation of identical models of opinion, vote and the covariance structure of the two surveys. While topline comparison of weighted and unweighted samples have played an important part in evaluating online samples, this comparison directly confronts the practical issue of whether the underlying multivariate structure of opinion is similar or different between online and live interview samples. We address this question by constructing multivariate models of opinion and vote for president and for Senate and conducting Wald tests for the equality of coefficients between the models. We also use model based estimates from each method to construct and compare predictive models based on ACS microdata of the Wisconsin population.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Reaching the Hard to Reach: Insights and Solutions

Getting to Know You: Strategies to Engage Hard-to-Reach Respondents

Bevin N Mory *Mathematica Policy Research*

Jillian Stein *Mathematica Policy Research*

Lisbeth Goble *Mathematica Policy Research*

Felicia Hurwitz *Mathematica Policy Research*

Conducting longitudinal survey research is particularly difficult among the young, at-risk population of the U.S. Department of Labor's YouthBuild program, which primarily serves high school drop outs. These challenges require innovative and adaptive methods to bolster survey response. During our third and final round of data collection, analysis of paradata indicated that the men in our sample, currently between the ages of 20 to 28, were underrepresented. To strengthen our understanding of this group, we conducted semi-structured interviews with YouthBuild Program staff who regularly engage with male program participants. These qualitative interviews provided insights into ways we can adapt our outreach to motivate young male sample members to respond. Among their recommendations, YouthBuild staff suggested sending a non-monetary incentive to encourage participation, arguing that this would be more enticing than the intangible promise of a monetary incentive upon survey completion. To test the effectiveness of this approach, we developed a random assignment experiment to investigate whether sending non-monetary incentives to young male nonrespondents increased survey response. The treatment group received incentive packages with a pair of colorful socks, a funny picture with a slogan (mimicking an internet meme), and an insert designed to encourage respondents to complete the survey online, which is the most cost-effective mode. To date, our experiment includes a sample size of 641 young males, and our preliminary findings suggest that sending targeted non-monetary incentives encourages participation and increases completion by web. This paper outlines the findings of our interviews with program staff, presents the results of the experiment, and discusses how other studies could apply this adaptive approach.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Reaching the Hard to Reach: Insights and Solutions

Evaluation of Cross-Survey Analysis Methods for the Estimation of Low Incidence Populations

Raquel Magidin de Kramer *Steinhardt Social Research Institute*

Elizabeth Tighe *Steinhardt Social Research Institute*

Population-based surveys are of limited utility to estimate rare, or low incidence groups, particularly those defined by religion or ethnicity not included in the U.S. Census. We have been using methods of cross-survey analysis, in particular, Bayesian Regression Modeling with Post-stratification to describe the geographic distribution and demographic composition of the Jewish population in the United States. With no external source of data to serve as a benchmark, questions regarding the validity of the method remain. The present study replicates the study using a sample of surveys in Canada, where the Census includes assessment of religious identification and we can compare population estimates derived from cross-survey analysis to the census. Three methods of cross-survey analysis are compared. These are: meta-analysis of complex surveys (MACS), pooled design-based cross-survey (PDCS), and Bayesian multilevel regression with post-stratification (BMRP). The accuracy and precision of these methods were assessed by comparing the estimates of the proportions of the adult Jewish population in the three largest Canadian metropolitan areas generated by each method with estimates of the Canadian census. The findings demonstrate that the BMRP method out-performed both the PDCS and MACS methods and yields reliable estimates of the population. This cross-survey approach provides a practical analytic framework that can be generalized both to the more extensive study of religion in the United States and to other social science problems in which single data sources are insufficient for reliable statistical inference.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Over Rated or Under Weighted? Methods for Improving Inferences from Online Nonprobability Samples

Going Beyond Geodemographic Weighting Adjustments to Reduce Bias in Nonprobability Sample Surveys

Robert Benford *GfK*

Frances Barlas *GfK*

John Lien *GfK*

Mansour Fahimi *GfK*

Elisa Chan *GfK*

With the growing interest in nonprobability sample surveys, during the past decade researchers have focused on sampling and weighting adjustment methods that can reduce some of the inherent biases of such surveys. Traditionally, bias reduction has been addressed through enforcement of geodemographic quotas and or weighting adjustments. However, recent findings suggest that inclusion of attitudinal/behavioral characteristics during weighting can further reduce bias of survey estimates and improve their external validity. For example, respondents from nonprobability online samples are more likely to be early adopters of new products than those coming from probability-based online panels. While it is important to control for basic demographics in nonprobability samples, it is unclear what set of non-demographic variables should be included as part of weighting adjustments. Nonetheless, as alternatives we use for survey administrations evolve, so must the methods we use for sampling and weighting adjustments. In this brief, a sampling alternative is presented that can improve the representation of general populations samples, where improvement is measured in terms of comparability of survey estimates to external benchmarks from government statistics.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Race, Religion, Sex and Gender

Investigating and Understanding Responses to Questions on Sex, Sexual Orientation and Sexual Identity

Carol Pierannunzi *Centers for Disease Control and Prevention*

William Garvin *Centers for Disease Control*

Machell Town *Centers for Disease Control and Prevention*

The Behavioral Risk Factor Surveillance System (BRFSS) conducts public health surveys in each US state and several territories. Beginning in 2014, states were provided the option to include a module on sexual orientation (straight, lesbian or gay, bisexual or other) and sexual identity (transgender male to female, transgender female to male, transgender nonconforming or not transgender). In 2014 and 2015, over 320,000 respondents from 27 states and one territory were asked questions from this optional portion of the survey. This paper will illustrate comparisons among responses to questions on sex from the demographic portion of the questionnaire, to responses on the module on sexual orientation and sexual identity. In some cases, responses on sex, sexual orientation and transgender identity are inconsistent. Such contradictions may illustrate that respondents are answering a common demographic question on sex from different perspectives. Patterns of responses are presented to show that traditional methods of asking respondents' sex may be biased and need refinement. In addition a review of characteristics of persons who refused to answer questions on the module was conducted. Responses to questions on sexual orientation and identity are also compared to other demographic characteristics (age, race, ethnicity, education) and whether the interview was conducted in English. There was evidence of associations between sexual orientation, identity and other demographics as well as associations between sexual orientation, identity and health related quality of life on the BRFSS. Using a large sample that is unlikely to be available from other sources, this paper provides information on question development to surveys which may be considering adding questions on sexual orientation and identity.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Whet Your Appetite - Survey Data Collection Using Smartphone Apps

The Feasibility of Using Smartphones to Record Food Purchase and Acquisition

Ting Yan *Westat*
Janice Machado *Westat*
Andrew Heller *Westat*
Aaron Maitland *NCHS*
John Kirlin *USDA*
Erika Bonilla *Westat*

The proportion of American adults owning a smartphone has almost doubled since 2011; now about two thirds of American adults own a smartphone of some kind and for many of them their smartphone becomes an important tool in their lives (Pew Research Center, 2015). Given this trend, researchers are also increasingly using smartphones for various research purposes such as ecological momentary assessment, transportation and time-use diary studies, health monitoring, and passive mobile data collection (e.g., Link, Lai, and Bristol, 2014; Sonck and Fernee, 2013; Revilla, Ochoa, and Loewe, 2016). This paper describes an additional use of smartphones to collect food purchase and acquisition. Food purchase and acquisition used to be collected through paper diaries. In a pilot study, we developed a smartphone APP that allows potential respondents to enter all foods and drinks they've obtained during a week on a smartphone. The APP also has functionalities that are supposed to reduce reporting burden such as the ability to allow respondents to scan barcodes, to use type-ahead to get food item descriptions, and to upload photos of foods and/or receipts. In this paper, we will discuss the feasibility of using this smartphone APP to record food purchase and acquisition in a national probability sample. We will focus on the proportion of respondents using this smartphone APP and their demographic characteristics and their experience with the smartphone APP. Implications of the findings will also be discussed.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Adaptive/Responsive Design: Predictions and Evaluations

Enabling Adaptive Design through Technology

Jerome Wernimont *Westat, Inc.*

Martha Stapleton *Westat*

Adaptive survey design includes an array of approaches for survey researchers to actively monitor survey progress while in the field, implement interventions based on incoming data, and assess the impact of the interventions on sample coverage and response rates. Most of these approaches rely on paradata and survey outcome data. Essential for any adaptive design effort is technology that efficiently orchestrates survey processes across data collection modes; collects detailed paradata; provides user-friendly feedback using data visualization; and seamlessly facilitates dynamic alterations to the survey design. The study design aspects that might be altered include data collection modes, timing of study activities, sample prioritization (e.g., propensity to respond), incentive amounts, and notification techniques and frequency. Altering any of these elements while in the field requires careful coordination, especially when multiple modes of data collection are being used. Where study support systems have not been developed to include adaptive features or characteristics, it is often not feasible to make some of the types of modifications due to system, resource, or schedule constraints. Integration technology can be employed to reduce these constraints, enabling study managers to incorporate more adaptive techniques. Integrated management of survey activities helps ensure the activities are executed as designed and to completion, and that the detailed paradata needed for feedback is available and consistent across modes. This presentation will discuss and provide examples of the use of Westat's M3 flexible integration and orchestration platform to support study operations while also efficiently enabling and managing adaptive techniques in studies that are in the field.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Measuring and Evaluating Nonresponse

Public Attitudes on Federal Statistics: What are respondents really thinking?

Gerson David Morales *U.S. Census Bureau*

Jenna Fulton *U.S. Census Bureau - Washington, DC*

Peter Miller *U.S. Census Bureau*

Jennifer Childs *U.S. Census Bureau*

Since 2012, the Census Bureau, collaborating with other federal agencies, has been tracking public views of the federal statistical system. The aim of this project is to understand the level of public trust in the measurements of societal conditions produced by the government's statistical agencies, and to see how trust is affected by external events such as cyber attacks. Monitoring trust could provide agencies with information needed to respond to events and to measure the relationship between trust and survey response rates. In its first two years, the tracking project, accomplished by adding questions to a Gallup nightly polling effort, appeared to show that trust in federal statistics can be negatively related to events, such as the 2013 federal government shutdown. However, causal inferences concerning the effect of such incidents have been limited because of the lack of survey measures of public knowledge of them. In 2014, the nightly survey was modified to include an open question, asking respondents to give reasons for their attitude reports. This open question – a “random probe” -- was programmed to randomly follow one of the several closed attitude questions measuring views of federal statistics. This paper reports on findings from an analysis of the open question responses in interviews conducted from 2014 to 2015. An interesting pattern emerges in the responses: those respondents who hold negative views of federal statistics more frequently ground their opinions in their views of the government per se, whereas those who hold positive views tend more to cite aspects of the statistical information. That is, distrust in federal statistics appears to be more rooted in antipathy for the government, while trust is based more on evaluations of statistical methods and products. We discuss the implications of these findings for government data collections and for measuring attitudes toward them.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Measuring and Evaluating Nonresponse

Nonresponse Trends in Telephone RDD Surveys

Nicholas A Hatley *Pew Research Center*

Courtney Kennedy *Pew Research Center*

Kyley McGeeney *Pew Research Center*

The Pew Research Center has conducted a number of studies aimed at measuring the impact of nonresponse on telephone surveys. Between 1997 and 2012 the response rate (AAPOR RR3) to the Center's telephone RDD surveys declined from 36% to 9%. Most estimates, however, did not show meaningful levels of nonresponse bias when compared to external benchmarks from federal surveys and other sources. This presentation reports the results of a 2016 Pew Research Center study updating this line of research. We update our response rate trend which, contrary to some media narratives, has plateaued around 9% rather than continuing to decline. We also compare trends in weighted telephone RDD estimates to trends from benchmark federal surveys such as ACS, CPS, and NHIS. If telephone RDD methodology is truly falling apart, we would expect to see increasing bias in the unweighted demographic composition of samples and increasing bias in weighted estimates over the past few decades. The data generally do not support that assertion. That said, telephone RDD surveys are not without their limitations. The civic engagement bias documented in previous Pew studies remains a serious challenge, and the 2016 presidential election raised questions about how well all public opinion surveys, including telephone RDD, represent Americans with lower levels of education and/or anti-establishment viewpoints.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Driving Them to the Web: Strategies, Techniques and Innovations

Too Good to Be True – Incentive Experiment Results from a Multi-Wave Student Survey

Karen Grigorian *NORC at the University of Chicago*

Jill Connelly *NORC at the University of Chicago*

College Point is a multi-wave evaluation of a virtual advising program offered to high-achieving, low-income high school students as they transition from high school to college. The evaluation tracks the effect of virtual advising on which colleges students apply, where they are accepted, and, ultimately, which colleges they decide to attend. The program and evaluation are funded by America Achieves. An incentive experiment was implemented during the first wave of data collection in spring 2016 when approximately 8,000 high school seniors who had been offered the opportunity for CollegePoint advising were contacted and recruited to complete the 10 minute online evaluation survey. As CollegePoint is a multi-wave data collection, the research team wanted to determine the optimal way to use incentives to gain cooperation in the base wave and sustain it in subsequent waves. The incentive was an Amazon gift code since these are easy to deliver and ideal for students, but the optimal amount, timing, and messaging associated with incentive was uncertain. To obtain reliable incentive information, we conducted an experiment that was designed to test three conditions: 1) Timing of incentive delivery (i.e., pre-paid versus post-paid); 2) Incentive amount (i.e., \$10 or \$15); and 3) Foreshadowing (i.e., foreshadowing additional incentives for future participation and a bonus for perfect participation or not foreshadowing). The survey sample was randomly divided into 9 experiment groups, each representing one combination of the three conditions listed above. We used serpentine sorting to maximize the similarity of adjacent cases during assignment. The group receiving a \$10 post-paid incentive without any foreshadowing served as our control group. This presentation focuses on the incentive experiment results and subsequent impact on the second wave of data collection.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Whet Your Appetite - Survey Data Collection Using Smartphone Apps

Challenges of Using an App to Collect Survey Data from Residents

Sonya Wytinck *National Research Center, Inc. - Boulder, CO*

Erin Caldwell *National Research Center, Inc.*

Apps hold promise for collecting survey data efficiently and in a way that appeals to people on the go. This seems especially enticing for a travel diary, where people are literally tracking where they go. NRC worked with the City of Boulder and DV-Mobile to develop an App to help collect data for the latest iteration of the City's travel diary study that has been conducted by NRC every three years for the past two decades. Through the results of a split sample we explore the viability of apps/smart phone use in resident surveys, including differences in respondent profile and travel behavior based on mode and potential impacts on differential response rates. Residents in this study were sent a set of instructions in which they were assigned a random day to complete the diary (tracking their trip types and travel modes) and asked to complete a short household survey to collect background demographics and transportation-related information. In the most recent iteration the sample was divided into three groups to test the use of the App: (1) received a letter with instructions to go to a website to download the App or download paper versions of the survey if they did not have a smartphone; (2) received paper versions of the instructions, travel diary and household survey with no reference to the App (similar to past iterations); and (3) received paper versions of all survey materials and also instructions to download the App if they would prefer to use it to complete the study. The standard method (no App option) had the highest response rates, followed by the combined method, with lowest response from those who were not mailed any paper survey materials.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Whet Your Appetite - Survey Data Collection Using Smartphone Apps

Smartphone GPS Applications as a Mode of Travel Survey Data Collection

Joann Lynch *RSG*

Michelle Lee *RSG*

Leah Flake *RSG*

Household travel surveys (HTS) are conducted in major metropolitan areas in the United States (as well as internationally). A typical HTS collects both demographic data and detailed travel diary data from all household members through a two-part survey process. The modes of data collection for an HTS were almost exclusively paper (mail-back) or telephone through the early/mid-2000's, at which time data collection became primarily web-based, while continuing to offer telephone response. Recently, some metropolitan planning organizations (MPOs) and state departments of transportation have adopted new technologies for HTS data collection using smartphone-based applications to collect both the survey data and the GPS travel diary data. By using a smartphone-based GPS application as the primary means of data collection for the majority of households (those that own smartphones), planning organizations can obtain more accurate and detailed spatial, temporal, and survey-specific data, particularly when compared to traditional data collection modes. Smartphone-based GPS applications also provide the opportunity to collect longitudinal and/or continuous data due to decreased respondent burden and administration costs. From 2014-2016, RSG conducted three comparable HTS efforts for North American MPOs including the City of Calgary (Alberta, Canada), the Puget Sound Region (Washington, USA), and the Triangle Region (North Carolina, USA). Across the three regions, a total of 11,800 households participated in a one-day travel diary either online or over the phone. For each MPO, a panel of households (approximately 6.4%) also completed a multi-day GPS study using RSG's smartphone-based GPS application, rMove™, on their personal smartphones in addition to the one-day diary. The panel results provide a direct comparison of traditional and smartphone GPS data collection modes at the household-, person-, and trip-levels. Data collection modes are analyzed based on the data quality, data quantity, respondent burden, and costs associated with each methodology.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Driving Them to the Web: Strategies, Techniques and Innovations

Who Can We Text? Assessing the Extent of Biases in Consent to Receive Text Message Reminders in a Follow-up Survey

Mengmeng Zhang *American Institutes for Research*

Rebecca Medway *American Institutes for Research*

Mark Masterton *American Institutes for Research*

As mobile phone adoption continues to grow and survey response rates continue to decline, sending text message to sample members is an attractive, low cost option for maximizing response. This is particularly true for younger populations, such as college students, where mobile phone adoption is nearly universal. However, the requirement to gain consent to text sample members before contacting them via text can be an impediment to its use. In addition, if text messages are successful at increasing response in subsequent surveys, and certain types of sample members are more likely to consent to receive them than others, then sending these additional texts may actually have the potential to bias results. In this presentation, we will report on the results of the baseline administration of a longitudinal web survey of college students in which students were asked to consent to receive text message reminders as part of a later, follow-up survey. Twenty-nine percent of students agreed to receive these messages. Using survey responses and frame data, we will assess whether there were biases in terms of who agreed to be contacted via text. We also will assess whether consenters were more willing to comply with other requests, such as confirming their email address and whether they provided higher quality data overall (e.g., less item nonresponse). Finally, using a texting experiment included in the follow-up survey, we will determine how successful the texts were at increasing response to the follow-up, and whether this increase was limited to certain subgroups of the population.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Driving Them to the Web: Strategies, Techniques and Innovations

Novelty of Text Messages as Reminders for Web Surveys: Does it last?

Kirti Kanitkar *The Gallup Organization*
Jennifer Marlar *The Gallup Organization*

Web surveys are increasingly being used as alternatives to traditional telephone and mail surveys, for reasons of cost, speed, and presumed appeal to younger demographics. However, web surveys suffer from low response rates, and just as in other modes, younger and minority respondents have lower response rates. Text message (SMS) communication has been recently explored as an effective alternative to email for improving web survey participation due to ease of access, given that 90% of Americans own a mobile device and sending and receiving text messages is the most common cellphone activity. A recent study indicated that text messages combined with email reminders were effective in improving the response rate of groups that had consented to receiving text messages, compared to a similar group who was only sent email reminders (Kanitkar & Marlar 2016). However, as with most modes of reminders – there is danger of a habituation effect, where, after the initial novelty of receiving SMS reminders wears off, SMS reminders produce very little or no increase in response rate above email reminders. The present study uses the Gallup Panel to explore the habituation of respondents to SMS reminders for web surveys. Panel members who provided consent to receive text messages were randomly assigned to receive SMS reminders or email reminders consistently across multiple surveys. Respondents in the SMS treatment received messages directing them to the web survey. The findings will explore differences in participation rates by reminder mode, as well as changes in propensity to participate by mode over time. Differences in participant demographics, device used to complete the survey (mobile versus computer), substantive answers, and timeliness of responses will also be explored. Panelist perceptions of text message reminders compared to email reminders once text message reminder become normative will be discussed.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Race, Religion, Sex and Gender

Gender Differences in Sources of Support for Gay Marriage

Claire Kelley *International Survey Center*
Sarah Kelley *UC Berkeley*

This paper explores the relationship between gender, attitudes about premarital sex, and the hotly disputed issue of gay marriage, demonstrating how attitudes about minorities can emerge from moral reasoning about more common practices. We find strong differences in the structure of attitudes about gay marriage between men and women. Not only are women substantially more supportive of gay marriage, net of religiosity and other background variables, but their support is also extremely strongly tied to attitudes about (heterosexual) premarital sex (standardized effect of .42), while it is much less so for men (standardized effect of .20). Additionally, while party identification drives Democrats to become slightly more supportive of gay marriage, there is a reciprocal effect for men with men (but not women) who oppose gay-marriage for other reasons also shifting to the Republican party. These patterns suggest underlying differences in the antecedents of attitudes about gay marriage for men and women: for women it is largely a “special case” of broader sexual morality, but for men it is something different (and often more threatening). Data comes from Round 20 of the International Social Science Survey, 2016 conducted in the United States with a representative national sample of 2295 cases. Analysis is by regression and structural equation modelling.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Over Rated or Under Weighted? Methods for Improving Inferences from Online Nonprobability Samples

Weighting and Estimation Procedures for Nonprobability Surveys: Variable Selection vs. Statistical Technique

Andrew W. Mercer *Pew Research Center*

By definition, nonprobability surveys cannot rely on random selection as a basis for statistical inferences. Instead, models are required to generalize from sample to population. For their models, researchers must choose the right set of covariates, correctly specify the relationship between those covariates and the outcome of interest, and replicate the population distribution of the adjustment covariates. While this is no small task for a single outcome, the situation becomes even more challenging when survey researchers are interested in many different outcome variables on a variety of topics, each of which may require different covariates or model specifications. Furthermore, the requirements may vary depending on sample source. At the same time, there is a desire for standardized techniques that work well out-of-the-box across a wide range of outcomes and sample sources. We present the results of a study by the Pew Research Center in which we compare alternative approaches to weighting and estimation for questions on a variety of topics pertaining to public opinion. We compare several different statistical techniques including raking, generalized regression estimation, matching, and multilevel-regression and post-stratification. We compare one-size-fits-all approaches to variable selection to those that are tailored for specific topics. These evaluations are performed using four large-scale nonprobability surveys from different vendors that were fielded in parallel by the Pew Research Center. We assess the advantages and disadvantages of alternative modeling procedures, the sensitivity of estimates to decisions about statistical technique and the selection of adjustment covariates, and the consistency of results across topics and survey providers.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Online Polls in the 2016 Election

The “Shy” Presidential Voter: Are voters more willing to be polled online (and tell us the truth)?

Debbie Ann Borie-Holtz *Rutgers University*

Ashley Koning *Eagleton Center for Public Interest Polling*

On the eve of the presidential election, randomly selected registered voters in four early voting states, namely Ohio, Florida, Nevada and North Carolina, revealed a divide was emerging between those who had cast early ballots and those heading to the polls just days before November 8th. A panel of voters randomly selected from the L2 registered voter database and recruited by cell to participate in an online survey signaled the lead held by Hillary Clinton was trending towards Donald Trump. Among those same voters surveyed, only 50 percent reported they had told a family member who they were voting for president. Just 42 percent had shared their vote preference with a friend and only a quarter of respondents had told a co-worker about their presidential decision. Clinton supporters were decidedly more willing to share their vote preference compared to Trump supporters by a 20 point margin. These preliminary findings suggest several possibilities: history had intervened in the final days before the election, voters were understating their likelihood to cast a ballot, respondents were unwilling to tell phone pollsters their preferences or all of these factors had converged. To examine these questions, we recruited voters to participate in a panel examining the 2016 presidential election volatility beginning with the pre-election weekend poll and on election day (N=728). Two additional panel polls are planned, along with an analysis to determine if the pre-election weekend trend that emerged was due to changes in voters' preferences, voter “shyness” disappearing online or in who actually cast ballots.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Online Polls in the 2016 Election

Measuring effect of Democratic voter crossover for Donald Trump through online voter networks

Aleks Mistratov *Brigade*

Matt Mahan *Brigade*

Almost all pre-election polls called for a resounding Hillary Clinton victory on November 8th, and yet, Donald Trump will be our country's 45th President. What happened and why did all the polls get it wrong? Brigade -- a mobile app and website that asks users who they'll vote for in the election and verifies their identity by matching them to their state's voter record -- discovered a trend missed by many pre-election polls. In the months leading into the election, Brigade saw a statistically significant ($R=0.8$) cross-over effect of registered Democrats pledging their votes for Donald Trump at disproportionately high rates in states like Pennsylvania ($N=2,895$) and North Carolina ($N=1,316$) -- states where Trump ultimately outperformed FiveThirtyEight election forecasts. Conversely, in states where Clinton outperformed pollsters' forecasts, like Nevada ($N=374$), registered Democrats on Brigade were significantly less likely to cross-over to pledge their vote for Trump. Could pollsters have picked up on this trend before Election Day and used it to improve election forecasts? While voters increasingly engage with politics online, their online political activity -- posts on social media, public endorsements of candidates, and answers to online polls -- are still under-utilized in election forecasts. Insights gathered from Brigade's pre-election data had the potential to improve 2016 Election Day forecasts, making Donald Trump's victory come as less of a surprise. This paper will outline Brigade's method of asking users who they pledge to vote for, its method for verifying users against their public voter record using information like name, date of birth, and mailing address, as well as a detailed analysis of registered Democrats in each state who pledged to vote for Donald Trump. It will also expand on how we believe data from voters' online activity can be used to aid future election forecasts.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Over Rated or Under Weighted? Methods for Improving Inferences from Online Nonprobability Samples

Using internet survey platform to sample online respondents

Jack Chen *SurveyMonkey*

Laura Wronski *SurveyMonkey*

Sarah Cho *SurveyMonkey*

Internet-based data collection is much criticized for the lack of a universal sampling frame in which selection probabilities are reliably determined. As a result, uncertainties around estimates from internet-based samples are often calculated based on single-stage sampling assumptions, which can greatly underestimate the underlying variation given the large internet-based sample sizes. This study explores different sampling schemes based on information collected on the SurveyMonkey platform, including device type, time of the day, day of the week, and types of surveys, to inform strategic sampling for internet-based surveys. The research utilizes over one million online respondents in the SurveyMonkey election tracking poll as the sampling universe. Variance estimation for multi-stage probability sampling design is extended to the internet-based sampling methodology. Results of the study lay the foundation for internet-based sampling and the theoretical framework for survey statistics from internet-based samples.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Over Rated or Under Weighted? Methods for Improving Inferences from Online Nonprobability Samples

Experimental Weighting Techniques for Online Nonprobability Election Polls

Jack Chen *SurveyMonkey*

Jon Cohen *SurveyMonkey*

Sarah Cho *SurveyMonkey*

For this study, we examine approximately 1 million registered voters to SurveyMonkey's Election Tracking poll from January to November 2016. Individuals were selected from a random sample of respondents to user-generated SurveyMonkey surveys with the opportunity to participate in our election poll. We performed traditional post-survey adjustment to weight the election sample to age, race, sex, education, geography, and voter registration status derived from Census data. To experiment weighting on factors outside of demographics, we apply machine learning techniques to determine the distribution of diverse survey topics conducted by more than 3 millions daily users of SurveyMonkey, and weight sample respondents' originating survey topics to SurveyMonkey platform topic distribution. In addition, we perform clustering techniques to group election respondents with respondents outside of Election Tracking poll based on device and browser types, and times of survey completion. These weighting techniques are still experimental, but they may form the basis for future polling done online.

May 20th, 2017
10:00 AM - 11:30 AM
Concurrent Session G

Online Polls in the 2016 Election

De-monopolization of Polls: What Do Ordinary Citizens Make of Different Quantifications?

Ozan Kuru University of Michigan

Whereas public opinion polls constitute a large part of the ways we quantify and disseminate mass opinion, due to more advanced and instant analytical/reporting tools in the digital age, the techniques to gauge, analyze, and report on the public's preferences have greatly diversified. Today, especially in the horserace context of elections, we have new tools ranging from polling aggregations/averages, forecasting models and google search-term analytics to social media analytics and election prediction markets. These relatively newer ways of quantifying the public opinion have distinct features such as being online, dynamic (instant-live-updating), integrative (diverse data incorporated), and more visual (interactive graphics). Whereas a substantial literature is developing to make sense of and measure how these new tools inform our understanding of public opinion (in relation to traditional self-report paradigm of polls), how the public makes sense and reacts to these new representations have not been examined. An exploratory survey (Study 1, N=200) and a survey-experiment (Study 2, N=400) are conducted in online samples to investigate how respondents make sense of, differentiate, and judge the credibility of these diverse tools in the context of 2016 U.S. presidential election. Study 1 examines how respondents compare and evaluate these diverse tools in general. Study 2 investigates the role of motivations in credibility evaluations and respondents' election predictions by presenting different types of hypothetical election reports where results would be either favorable or unfavorable to the respondents. Results reveal that important differences exist in people's exposure, familiarity, and evaluations of these tools. However, these differences are mostly trumped by the motivational biases, suggesting that these "more enhanced" tools of tracking public opinion might not necessarily be perceived as "stronger" evidence or corrective against motivational resistance. The implications for the meaning of public opinion in digital age, public perceptions of public opinion, and public opinion research itself are discussed.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Raking and Weighting ANES Time Series

Joseph Wu *Stanford University*
Jon Krosnick *Stanford University*
Matthew DeBell *Stanford Univ.*

The ANES (American National Election Studies) Time Series surveys have been conducted every 2 years from 1948 to 2008, using a combination of fresh cross sections and panels, including a pre-election and a post-election component each survey. Many of these survey datasets are unweighted. In this project, we are building one of the most comprehensive and high-quality sets of weights for every ANES Times Series dataset dating from 1948 to 2008. For each ANES year, we build base weights from the ANES dataset to account for probability of selection, defaulting to the NEHM (number of eligible household members), adding adjustments for (if needed) panel attrition and sampling bias. Then, demographic and election variables are processed and raked with our base weights, through a post-stratification weighting methodology described in “Computing Weights for American National Election Study Survey Data” by DeBell and Krosnick (2009), implemented through the ‘anesrake’ R package, maintained by Josh Pasek. The final raked weights, for pre-election and post-election, are tested against CPS (Current Population Survey) benchmarks as well as election results. The R code, final weights, and result comparison tables are maintained for each ANES year. Raked variables achieve much higher accuracy to the benchmarks while unraked variables have also shown improved accuracy over the unraked data.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

The Transition from Landline to Dual Frame RDD Surveys: Findings from the Asahi Shimbun

Yasuyuki Saito *The Asahi Shimbun*

Nicolaos E. Synodinos *Univ. of Hawaii - Manoa*

Japanese media have been using landline RDD samples for public opinion surveys since around 2000. Recently, concerns increased because of the decreasing use of landline phones and the difficulty of reaching younger respondents with landline frames. The latest estimates of Asahi Shimbun indicate that in Japan about 15% live in households without a landline. In terms of usage, 25% are cell phone only (CPO) and another 36% are cell phone mostly (CPM). The CPO are almost 90% among singles in their 20s and 30s. In the last landline survey (July 2–3, 2016), 9% of the respondents were under 40 as compared to 29% in the Census. In 2014, two wireless sampling frame pilot surveys were conducted by a consortium of six Japanese media companies. In addition, Asahi conducted seven pilot surveys using wireless frames during 2015–2016. These surveys were conducted concurrently with Asahi's landline RDD surveys. Their purpose was to refine the field operation methods and the weighting for the dual frame surveys that started in mid-July 2016. The latest (October 2016) 2-day survey had 1,000 landline and 1,036 wireless respondents. For telephone numbers with a Japanese adult, 53% responded to landline calls and 49% to wireless calls. Only 3% of respondents were 18–29 in the landline frame compared to 14% in the wireless frame. Conversely, 35% of those in the landline frame were over 69 compared to 12% for the wireless frame. Landline and wireless respondents replied similarly in most opinion questions. However, there were some large differences including reported voting behavior and views towards the Prime Minister's economic policy. Lessons learned in conducting and weighting dual frame surveys in the Japanese context will be discussed and comparisons will be presented for respondent characteristics and opinions between the landline and wireless frames.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Nonresponse in a National Survey of Board Certified Nurses

Luciano Viera *American Nurses Credentialing Center*

Chie Ohba *American Nurses Credentialing Center*

The American Nurses Credentialing Center (ANCC), a subsidiary of the American Nurses Association (ANA), is the world's largest nurse credentialing organization. Each ANCC specialty practice area certification exam requires a periodic job analysis that includes a national survey of board-certified nursing populations. Nonresponse in surveys of nurses, the largest population of healthcare professionals, threatens the validity of such healthcare research and assessment efforts. In this study, available sample frame data (e.g., certification renewal status, race-ethnicity, gender, and geography) will be examined to determine if they are related to survey nonresponse. Preliminary results suggest that higher response rates were associated with nursing certification renewal. Implications for surveying healthcare professionals will be discussed.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Augmenting Survey Results with Geographic Information

Emily Goettsche *Westat*
Emily Goettsche *Westat*
April Oh *NCI*
Michael Giangrande *Westat*
Laura Dwyer *NCI*
Terisa Davis *Westat*
Dave Stinchcomb *Westat*

This presentation will discuss how survey research can benefit from the addition of information about neighborhood context based on geographic location. There is a growing interest in neighborhood context in health studies. Research has shown that neighborhood characteristics including poverty, unemployment, and crime are key determinants of an individual's health. In this talk, we show how results from the FLASHE (Family Life, Activity, Sun, Health, and Eating) Study sponsored by the National Cancer Institute (NCI) were augmented with data about neighborhood characteristics. These methods can be applied to any type of survey where information about neighborhood characteristics can add research value to the results. FLASHE is a study that examined the psychosocial, generational (parent-teen), and environmental correlates of cancer-preventive behaviors. During data collection, caregivers and their adolescent children were asked similar questions about diet and physical activity behaviors. NCI asked Westat to augment the survey data with neighborhood context information based on geographic location to allow researchers to explore the relationship between survey results and neighborhood factors. We had access to both the location of the home and the location of the teen's school. We created a set of buffers of varying size and added contextual information using data from the U.S. Census Bureau. We added demographic variables including population density, percent of population by race, and urban/rural category. We also added a set of socio-economic characteristics such as median household income and percent unemployment and then created a summary measure of neighborhood inequality, the SES index, using these variables. Finally, walkability factors were created to summarize measures of neighborhood walkability using principal component analysis and built environment measures. This presentation will show how geographic information supplementation is an inexpensive and effective way to increase the utility of survey data.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Selecting appropriate time points for trend analyses of the National Health and Nutrition Examination Survey

Jennifer R Rammon *National Center for Health Statistics*

Jennifer Parker *National Center for Health Statistics*

Deanna Kruszon-Moran *National Center for Health Statistics*

Data from the National Center for Health Statistics' (NCHS) National Health and Nutrition Examination Survey (NHANES) are often used to detect the extent to which various health problems and risk factors have changed in the U.S. population over time. Trends in obesity, undiagnosed diabetes, blood lead levels, and diet have been used by other government agencies and private sector organizations to establish policies, plan future research, and develop health promotion programs. Since NHANES data is released in two year cycles, time points in trend analyses are often presented at two year intervals. For some analyses, however, one or four year intervals may be preferred. Based on the NHANES sampling design, at least four years of data is required to obtain acceptable levels of reliability for many estimates representing subdomains defined by sex, race, and age. Conversely, in contrast to one year intervals, two or four year intervals could theoretically increase the variance of the slope estimates. This project examines the differences between using time points that are defined by publicly available two or four year time points versus those that are defined by limited access one year time points. Preliminary results examining systolic and diastolic blood pressure in adults show that the three methods are comparable when evaluating a strict linear trend. Neither the slope nor the variance estimates changed when choosing two or four year time points over one year time points. Further work will examine a larger variety of health outcomes with different trend patterns to determine if this comparability remains for nonlinear trends among the major subdomains defined by gender, age, and race/ethnicity for which NHANES was designed to analyze.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Predicting Response and Ethnicity Status Among Minorities in Random Digit-Dial (RDD) Surveys

Michael Jacobsen *RTI International*

Rebecca Powell *RTI International*

Matt Jans *UCLA Center for Health Policy*

Hard-to-reach populations, such as racial, ethnic, and linguistic minorities, are challenges to the representation of all surveys. They are also an operational challenge for surveys in which the target population includes a large proportion of hard-to-reach respondents. Accurately predicting the number of racial or ethnic subgroup respondents in such surveys prior to contacting households could reduce both coverage bias and survey costs by assisting in targeting minorities. RDD surveys present unique challenges to such prediction, as RDD samples, especially cell phone frames, tend to have little auxiliary data limiting the coverage and ultimate accuracy in both minority status and response prediction. Augmenting the sampling frame with other forms of data should significantly improve response prediction models (Peytchev et al., 2009). Using this method of augmenting the sampling frame, we test the ability to predict the number of respondents (i.e., completed interviews) obtained from racial and ethnic subgroups in the California Health Interview Survey. Specifically, this research develops a nonresponse prediction model using sampling frame data, computer-assisted telephone interview (CATI) paradata, and publicly available demographic data from the U.S. Census Bureau to attempt to predict the number of minority respondents. Poisson models using the above data are created to simulate the outcome of third quarter of data collection for the 2016 California Health Interview Survey and the correlation between the simulated and actual results are examined to evaluate the predictability of the model. Models with high predictability provide a framework for more accurately predicting subpopulation members that can be reached by telephone interviewers, thereby reducing potential coverage error and improving call center efficiency.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Within-Household Respondent Selection and Total Survey Error Trade-offs

Andrew T Dyer *ICF International*
Randal ZuWallack *ICF International*
Rob Andrews *NOAA*
James Dayton *ICF International*

The National Oceanic and Atmospheric Administration (NOAA) collects data through the Coastal Household Telephone Survey (CHTS), a landline, random digit dial telephone survey, to estimate marine recreational fishing effort in U.S. coastal counties. In the current design, an informant (i.e., the person who answers the telephone) answers screening questions on fishing effort for the household to determine whether or not a household is classified as a fishing household. To test the impact that different within-household sampling procedures have on survey measures, records were assigned to one of two treatment groups: a control group where questions were administered to any adult that answered the telephone, and an experimental group where questions were administered to one adult randomly selected via the Rizzo-Brick-Park method to report for the household. In addition to testing for an effect on survey measures, the analysis will take a Total Survey Error approach that includes examining paradata and the attendant effects of within-household selection on survey response, sample characteristics, demographic representativeness, and operational outcomes.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Do We Need the Stragglers?

Karen Jaffe *Consumer Reports*

Karen Jaffe *Consumer Reports*

Simon Slater *Consumer Reports*

Steven Witten *Consumer Reports*

There is mixed evidence regarding whether early and late responders to surveys differ demographically, psychographically, or concerning shopping habits and preferences. When looking at phone and paper mail-in surveys, some have found early and late responders to be virtually the same (Siemiatyk & Campbell 1983). On the other hand, studies looking at alcohol consumption found late responders are more representative of non-responders and, therefore, can act as a proxy for the group that introduces error into all surveys via their inaction (Meiklejohn, Conner, Kypri 2012). In this poster, we aim to add to this body of research by looking at how people who enter a web-based survey via a late reminder (reminder 4 or 5) differ from those who respond to the initial invite or the first or second reminder. To answer this question, we analyzed results of our 2016 Annual Summer Survey, where about 275,000 Consumer Reports Subscribers answered at least one section of a 12-section omnibus that asks questions ranging from appliance reliability to online dating habits and services. Preliminary analyses suggest there are not significant differences on a variety of psychographic questions gauging factors such as confidence, skepticism, and brand loyalty between those who enter a survey when first invited versus a late reminder. In addition, our analyses do not show demographic differences based on when people enter and finish a survey. These findings might suggest that we do not need late-wave respondents. However, half of all respondents came in during our reminder campaign, and over 25,500 people answered at least one section based on reminders four and five. We will discuss the pros and cons of sending several reminders, specifically exploring if the time and money spent are worth the higher number of respondents.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Are family caregivers hard to reach? Comparing approaches to maximize survey response

Tammy J. Payton *National Marrow Donor Program/Be The Match*

Christa Meyer *National Marrow Donor Program/Be The Match*

Kate Houg *National Marrow Donor Program/Be The Match*

Jenna Umar *National Marrow Donor Program/Be The Match*

Heather K Moore *National Marrow Donor Program/Be The Match*

Jill Randall *National Marrow Donor Program/Be The Match*

Ellen M Denzen *National Marrow Donor Program/Be The Match*

Blood and marrow transplant is an intensive, potentially curative therapy for blood cancers, genetic and immune disorders. Most hospitals require patients to have a full-time caregiver, often a spouse or family member, for an extended period of time after transplant. Caregivers are responsible for supporting the patient's physical and emotional health throughout the transplant process. Caregivers may not feel there is enough time to reach out to others for help or support, making them "hard to reach". The Caregiver's Companion Program (CCP) was developed to provide caregivers with self-care tools to help them cope with their everyday challenges and demands. The CCP is led by experienced coaches who provide one-to-one telephone support and guidance on how to use the program "toolkit". A satisfaction survey was administered to participants (N=116) to evaluate program effectiveness from November 2014 to present date. To maximize response rates, we compared two approaches: U.S. mail (Group 1) versus phone/email mixed-mode (Group 2). Program participants were assigned to Group 1 or 2 based on registration date. Group 1 received a cover letter and survey with a follow-up letter and survey two weeks later. For Group 2, an initial phone call was made after completing the program and second attempt one week later; invitation and survey link were sent by email one week later. A 17% higher response rate was achieved in the mixed-mode group (54% v. 71%). In addition, richer narrative was captured for open-ended questions during the phone interviews. Despite time and energy constraints due to their caregiving role, participants were willing to provide feedback on program effectiveness. Introducing a phone option made it feasible and potentially more convenient for those who are in and out of the hospital, potentially away from home for extended periods of time, to respond and elicit robust qualitative data.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Impute, Model, Select - A Model Driven PPS Sampling Method using Imputed Commercial Data

Paul W Burton *Institute for Social Research - University of Michigan*

Rare and difficult to locate populations are increasingly the target populations of interest in public opinion research. The naturally occurring low eligibility rate of these populations can lead to overly large sample sizes and high screening costs for the study. Our goal with this project is to significantly increase the screen-in eligibility rate for households with at least one 3-10 year old child in order to run a more efficient screening effort. In order to do this, we take advantage of the relationships between aggregate Census data and commercially purchased address level data to impute demographic measures for the entire data. At the first stage, Census data is used to determine Block-Group level eligibility propensities which are then used as the measure of size in a proportional sampling routine. The second stage consists of imputing demographic measures for the entirety of the data and then modeling eligibility propensities at the address level. Addresses are selected at the second stage proportional to this modeled eligibility propensity. Using simulations, we estimate that this procedure will increase our field screening eligibility by 50%. The method is currently being tested in the field and the results of the field work will be available prior to the conference.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Predicting Postmaster Returns in Nebraska

Patrick Habecker *University of Nebraska-Lincoln*

Alian Kasabian *SBSRC at Univ of Nebraska-Lincoln*

Address-based sampling is a dominant mode of survey research in the United States due to low cost and high coverage rates. However, little is known or has been published about variation in postmaster returns (mailings which are returned as ineligible by the postal service). Nonrandom postmaster returns have implications for cost projections, survey design, and survey adjustment. Common delivery sequence file frame information such as record type (e.g. street), delivery type (e.g. curb), level of rurality, and carrier route type can provide insight into understanding postmaster returns. We use multilevel models where addresses ($n=3467$) are nested within postal carriers ($m=1421$) to explore these associations. Postal carriers may be either a city delivery or rural route; although both types deliver across all spectrums of rurality. Overall, 9% of the addresses in our sample had a postmaster return. Preliminary analyses show rural route carriers had more variance in postmaster returns across rurality classifications than city delivery carriers. Where city delivery carriers were consistent, rural route carriers had higher levels of postmaster returns in rural areas, but lower levels in urban areas. In smaller towns, city delivery and rural route carriers were indistinguishable. Delivery type and record type are also significantly associated with postmaster returns. We find that there is a significant association between both address characteristics and carrier characteristics with postmaster returns in Nebraska. Implications for cost projections, survey design, and survey adjustment for address-based sampling surveys are discussed.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Do You Agree or Strongly Agree with the Following Statement? Data Visualization is an Important Part of Public Opinion Research: Strongly Agree!

Nola du Toit *NORC at the University of Chicago*

Edward Mulrow *NORC at the University of Chicago*

Naomi B. Robbins *NBR-Graphs*

The purpose of this poster is to demonstrate the everyday application of data visualization in public opinion research. Data visualization is a rapidly growing area of application for analysis and dissemination of data. It is used across a variety of fields, including biology, social sciences, and statistics, and can be applied to both quantitative and qualitative data. Data visualization as exploration can be used to examine trends in data, isolate discrepancies, and identify patterns that would otherwise be overlooked. Data visualization can also be used for explanation or presentation of research findings. Discoveries can be presented in an easily digestible form to a variety of audiences and across numerous platforms. We will present recommended methods for extracting interesting and useful findings from data using various tools and techniques. Moreover, we will discuss using data visualization to disseminate our findings to a wide range of audiences.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

School Climate Survey Compendium

Pia Peltola American Institutes for Research

In 2015, the No Child Left Behind legislation was replaced by the Every Student Succeeds Act (ESSA), which requires states that receive Title I funds to collect data on at least five indicators, one of which must be a measure about “school quality or student success,” such as a measure about school climate or social emotional learning skills. As schools prepare to implement the new law, there will be an increased need for surveys that measure school climate and learning conditions. The demand for school climate surveys have also increased as schools attempt to reform their learning conditions and monitor progress on an ongoing basis. This poster presentation introduces researchers to a School Climate Survey Compendium (SCSC), which is a website sponsored by the U.S. Department of Education’s National Center on Safe Supportive Learning Environments (NCSSLE) and maintained by staff at American Institute for Research <https://safesupportivelearning.ed.gov/topic-research/school-climate-measurement/school-climate-survey-compedium>. The SCSC is an invaluable resource to survey researchers because it includes a collection of screened surveys, assessments, and scales related to school climate that can assist researchers in their efforts to identify measures to assess school conditions. The scales in the surveys have been tested for validity and reliability. The compendium, which continues to grow, includes currently 22 student, 18 faculty and staff, and 9 family surveys at pre-K, elementary, middle, high school, and higher education levels. During the poster session, the presenter will provide information and answer questions about the surveys, how they have been screened, who conducted them, what types of constructs and measures are included in the surveys, where one can find reports and additional information about the surveys, and how one can nominate a survey to be included in the compendium.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Effects of Initial Contact Letter Content and Enclosures on Survey Response Rate

Jennifer M Renner *HealthPartners Institute*

Jeffrey P. Anderson *HealthPartners Institute*

Abigail S. Katz *HealthPartners Institute*

Jeanette Y. Ziegenfuss *HealthPartners Institute for Education and Research*

The prenotification mailing can establish legitimacy of a survey while appealing to a respondent's motivation to help others (i.e. social exchange theory). The ability or inability to leverage this social exchange can impact a study's response rate. It has been proposed that the tone of the initial contact letter may influence a subject's willingness to respond to the survey. Another detail that may impact response is the inclusion of an item of perceived value in the prenotification mailing. Our study, which used a survey to evaluate health education materials for women, explored these ideas by using cover letters with two different appeals, and by including or excluding a brochure in the initial mailing. The altruistic cover letter appeal used social exchange theory by asking subjects to "help improve women's health" and gave respondents the opportunity to "help HealthPartners and the Minnesota Department of Health improve health education materials." The egotistic appeal told subjects that "we need [their] input!" and that responding to our survey would give them "the opportunity to be heard." Additionally, a subset of the sample was asked to evaluate and respond to a survey on an enclosed brochure, while another subset was asked to evaluate a website, so did not receive a brochure. Our findings show that neither the appeal of the cover letter nor the inclusion of a "valuable" brochure impacted the response rate, suggesting that we were not able to successfully leverage for increased likelihood of response.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Exploring the relationships among days to complete, incentives, and response rates: A Case Study of 18-Month & 40-Month SaveUSA Surveys

Scott Peecksen *Decision Information Resources, Inc.*

Ronald McCowan *Decision Information Resources, Inc.*

Sylvia Epps *Decision Information Resources, Inc.*

Ronald Bass *Approximetrics*

Jo Anna Hunter *MDRC*

Achieving sufficient survey response rates in a cost-effective manner with longitudinal and hard-to-reach samples continues to be difficult. Researchers have partially addressed this difficulty by offering incentives that have improved response rates (Singer, Groves, and Corning, 1999). Less is known however about how “days to complete a survey” (DTC) may influence response rates among hard-to-reach populations. Only a few studies have examined predictors of DTC (Dinglas et. al, 2015) or used DTC to predict response rates (McCowan and Epps, 2015). McCowan and Epps (2015) found that higher survey incentives at wave 1 predicted fewer DTC at wave 2 and that wave 1 DTC predicted response rates at wave 2. Other studies have found that age predicted survey completion time (Dinglas et. al, 2015; Tenant and Badley, 1991) and response rates (Natasha et al., 2013). Using paradata, this poster will build on these findings by examining age, incentives, and gender as predictors of DTC and DTC as a predictor of response rates among a hard-to-reach population. The following research questions will be addressed: • Do age and gender predict DTC at wave 2 while controlling for incentive amount received at wave 1? • Do DTC at wave 1 predict response rates at wave 2 while controlling for age, gender, and incentive amount at wave 1? Results may inform the development of new incentive or other strategies that enhance response rates among hard to reach populations by decreasing respondents’ DTC. Data are from the Social Innovation Fund SaveUSA evaluation, a randomized control trial designed to test the effects of a savings program on various outcomes among a hard-to-reach population.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

The End of the Line for Landlines?

Ashley H White *University of Oklahoma Health Sciences Center*

Calls to cell phones tend to be more expensive than calls to landlines in Random Digit Dialing (RDD) phone surveys. Reasons include the higher cost of cell phone sample and the time required to manually dial cell phone numbers. A study by Pew Research Center found that demographics of cell phone respondents were more representative of the U.S. than respondents contacted on a landline and found few differences in poll results between the samples. We investigated differences in cost and key outcomes and demographics between cell phone and landline respondents to a survey conducted in Oklahoma. A health-related survey of adults aged 18 years and older was conducted between 2/2016-5/2016 (n=737 cell, n=265 landline respondents). Responses were weighted and raking was used to derive estimates reflective of the Oklahoma adult population. We found that the cost ratio is reversed, with completions to cell phones costing about ½ of landline completions, due to higher response rates (13% vs. 6%, AAPOR RR1). This offset the higher cost of sample and labor cost to manually dial. There was no significant difference in the percentage of current smokers among respondents from the cell phone compared to landline samples (23% vs. 17%, p=0.1033). Demographics from the two groups did not differ significantly by race, gender, education, or income, but did differ by age. There was no significant difference in the percentage of registered voters between cell and landline respondents (79% vs. 85%, p=0.1090), but a significantly lower percentage of cell phone respondents said they had voted in the last year (64% vs. 86%, p<.0001). Because of these findings, the cell phone to landline ratio for dual RDD surveys should be reevaluated. Future research will compare responses to other health-related questions both for general population surveys and those with special target populations. Funding source NIGMS 1 U54GM104938

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Do Sequential Mixed-Mode Surveys Reduce Nonresponse Bias and Measurement Error? An Experimental Study

Joseph Sakshaug *University of Manchester*

Alexandru Cernat *University of Manchester*

Trivellore Raghunathan *University of Michigan*

Mixing multiple modes of survey data collection is becoming a standard practice in survey research. Mixed-mode surveys are faced with a slew of design decisions regarding which types of modes to administer and which sequence to administer them in. Such decisions are largely based on administrative objectives, such as minimizing costs and maximizing response rates. However, just as important to these mixed-mode decisions is their impact on data quality, namely, nonresponse and measurement error, which is an understudied issue in the mixed-mode literature. In this research note, we report on a sequential mixed-mode experiment of young adult drivers randomized to one of two mode sequences: an interviewer-administered (telephone) mode with self-administered (mail) follow-up, or the opposite sequence. We address two open research questions: 1) Does implementing a second mode of data collection result in better data quality (i.e., less nonresponse and measurement error) compared to the starting mode? And 2) does the mode order of the sequential design impact data quality? By utilizing administrative records for the full sample, we find that introducing a second mode improves data quality compared to the starting mode, irrespective of which mode sequence is used. The analysis also indicates that the sequence starting with the self-administered mode results in better data quality compared to starting with the interviewer-administered mode.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Assessment of Non-Response Bias in the National Health and Nutrition Examination Surveys

Tala Fakhouri *The National Center for Health Statistics*

Joseph Afful *Harris Corporation*

James Dahlhamer *The National Center for Health Statistics*

Jennifer Parker *The National Center for Health Statistics*

The National Health and Nutrition Survey (NHANES) is a nationally representative survey of the health and nutritional status of the U.S. civilian non-institutionalized population. Like other face-to-face surveys, NHANES is being challenged by declining response rates in recent years. However, it is unclear if this decline in response rates has increased nonresponse bias in survey estimates. We propose to examine the impact of the declining response rates on bias by comparing weighted NHANES estimates for health outcomes and health-related behaviors to the same outcomes for the same time period from a larger face-to-face survey, the National Health Interview Survey (NHIS). In addition, we will examine changes in estimates from NHANES 2001-2004 to NHANES 2011-2014 directly and by comparing the corresponding change in the NHIS. We plan to do these assessments overall and for subgroups defined by age group, race and ethnicity, and socio-economic status. Findings from this analysis may help shed light on the impact of declining response rates on nonresponse bias in NHANES.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Who Lives Here? Inconsistent Reporting in the National Household Education Survey Household Enumeration

HarmoniJoie Noel *American Institutes for Research*

Danielle Battle *American Institutes for Research*

Carol Wan *American Institutes for Research*

Mahi Megra *American Institutes for Research*

Household enumeration and within household screening are important for identifying the appropriate individual to sample in household surveys, but sometimes there is reporting error around who lives in the household that can lead to coverage error. This paper will present results from a study evaluating inconsistent reporting in the household enumeration questions used in the National Household Education Survey (NHES). NHES enumerates households twice: 1) in the screener survey and 2) in the topical survey completed after screening. The same question is used in both surveys, yet in prior rounds of mail-based data collections about a quarter of households report a different total number of household members in the screener than in the topical survey. Usually, fewer household members are reported in the screener, which is problematic for within household screening. However, with up to a two month lag between the screener and the topical surveys, real changes in the household composition could have occurred during that lag. In 2016, a web mode which allowed the screener and topical surveys to be completed in the same sitting was piloted, reducing the likelihood that discrepancies could be caused by true differences in household composition. This paper will compare the percentage of households where there was a discrepancy between the total household size reported in the screener and the topical survey between the 2016 web and mail modes. We will use a logistic regression to analyze predictors of inconsistent reporting such as household size, survey mode, time lag between screener and topical surveys, respondent consistency from screener to topical, whether the inconsistent reporting was for an adult or child in the household, and demographics of the respondent. The presentation will conclude with a set of recommendations that can be applied to NHES as well as other household enumerations.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Quality of Commercial Data Sources

Antonia Warren *Westat*
Shelley Brock-Roth *Westat*

Survey researchers are facing more and more challenges in the pursuit of high quality and cost effective data. One critical challenge has been declining response rates which have led researchers to explore using commercial data when adjusting for nonresponse or to inform data collection efforts (West et. al 2015). These databases can provide a variety of auxiliary variables including demographic characteristics, geographic, political, and lifestyle information. Given the possibilities, researchers have started to assess the quality of these data and have found that certain variables can suffer from high rates of missingness or error (West et. al 2015). More methodological studies are needed to assess the quality of data from different vendors. The purpose of this study is to evaluate the ability of two vendors, Aristotle and Acxiom, to accurately identify current members of the household and assess the amount of missing data in demographic items. In this study, roughly 1,000 addresses from respondents in two surveys were compiled: the Survey of Free Time and the Health Activities and Lifestyle Survey. The vendors used the addresses to append auxiliary data for every person living at a particular address. Descriptive analyses will be used to evaluate the demographic items and a matching algorithm will be developed to assess whether each vendor has identified members of the household as reported by survey respondents. The results of the latter analysis could have implications for within household selection of a respondent or pre-data collection strategies, such as mailing an invitation letter to a specific household member.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Increasing Trust in Social Media Research through Improving Disclosure Standards

Ganna (Anna) Kostygina *NORC at the University of Chicago*

Yang Xu *NORC at the University of Chicago*

Yoonsang Kim *NORC*

Sherry Emery *NORC at the University of Chicago*

While the number of studies using “big data” is growing rapidly, few transparently outline their data collection and analysis methods. Without such reporting, it is impossible to evaluate the validity of findings or compare results across studies. For instance, social media platforms represent novel and rich sources of observational “big data” for health and public opinion research. As with any data source, how the data are obtained shapes the conclusions that can be drawn. Each social media platform has different technical constraints and poses unique methodological and programming challenges that define key parameters of data collection, nonetheless common disclosure principles can apply. We aim to develop guidelines for social media research minimal disclosure standards to report essential information about the research methods, and thus to fulfill the obligation for good professional practice and increase transparency, comparability, and replicability of social media research and analysis. Namely, the objective is three-fold: (1) to build a conceptual framework for social media data collection and quality assessment, (2) to propose a reporting standard for social media research, and (3) to apply this framework using real-world examples from tobacco-related Instagram posts. For the purpose of the study, we synthesize literature from biostatistics, computer science and health communication to develop a conceptual framework for standardized reporting of social media data collection and analyses. We then apply the framework to a specific example using approximately 2,000,000 Instagram posts. Thus, we estimate retrieval precision (how much of retrieved data is relevant) and retrieval recall (how much of the relevant data is retrieved) of tobacco-related content on Instagram, and show how choices of sampling and search filters affect results and conclusions. We propose that the disclosure standards include description of search filter assessment, data verification procedures, human coding sampling frames. Proposed framework presents an effective way for quality evaluation of social data.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Influence of Survey Administration Mode on Reported Opinions on Disparities in Healthcare Access in a National Multi-Mode Survey

Rachel Bavley *NORC at the University of Chicago*
Heather Morrison *NORC at the University of Chicago*
Ned English *NORC at the University of Chicago*
Kathleen Santos *NORC at the University of Chicago*

A substantial body of literature has demonstrated that the mode of survey administration can influence response, with interviewer-administered modes resulting in higher rates of socially desirable responses compared with self-administered modes. Moreover, response option order has been found to influence response tendencies differentially during visual vs. aural administration. Our research studies the effect of survey mode on items related to politically-sensitive subjects such as perceived disparities in healthcare access among minority or disadvantaged groups. Respondents to a national multi-mode survey were given the option to complete the survey by web, mail, or phone. All respondents were asked whether they thought certain groups (African Americans, Hispanic Americans, and low-income Americans) found it easier, harder, or about the same as more privileged groups (White Americans, Americans better off financially) to get the healthcare they needed. The effect of survey mode on responses was analyzed using multinomial logistic regression. Respondents who completed the survey by web or mail were significantly more likely to say that it was easier for disadvantaged groups to obtain healthcare, while respondents who completed by phone were more likely to say there was no difference in healthcare access between groups. While selection bias may have influenced these results as mode was not randomly assigned, these effects were independent of demographic factors such as gender, age, race/ethnicity, and education level. This presentation will explore the implications of these mode effects for healthcare and health disparities research. Our results are useful to practitioners of both single and multi-mode surveys as well as those who study sensitive topics.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Evaluating Sample Members with Ported Telephone Numbers

Timothy J Nesius *RTI International*

Amang Sukasih *RTI International*

Jessica Williams *RTI International*

Tamara Terry *RTI International*

Sample management between landline and cellphone telephone numbers is a critical element of operational success for Computer Assisted Telephone Interviewing projects. As part of standard sampling procedures telephone numbers are identified as either landline or cellphone numbers. Sponsored by the Centers for Disease Control and Prevention, the National Intimate Partner and Sexual Violence Survey (NISVS) is a national telephone survey of men and women drawn from a list-assisted RDD landline sample and an RDD cell phone sample. For the landline frame during the 2016 launch of the project, we examined a portion of telephone numbers screened via our TCPA compliance process and found approximately 700 cases were reported as having been ported from landline telephone numbers to cellphone telephone numbers. This group of landline-to-cellphone cases were sent an introductory letter since they had been included in the original landline sample whereas the existing cellphone sample did not receive an introductory letter. Identifying the group of landline-to-cellphone cases presents us with a unique opportunity to examine how this subset of cases performs. Our presentation seeks to compare the performance of the landline-to-cell phone cases that received a letter to the original cellphone sample. We will also examine the demographic differences among respondents with ported numbers, test the vendor flags indicating landline versus cellphone, and verify the accuracy of the addresses associated with the ported cell phone cases.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Behind the 2016 Election: Shifting Primary Preferences and their Implications in the General Election

Hannah Fingerhut *Pew Research Center*

Bradley Jones *Pew Research Center*

This poster will analyze individual-level changes in candidate preference in both the Republican and Democratic primary elections. The analysis will provide a window into change and stability in voters' choices over the course of ten surveys conducted throughout 2015 and 2016 on the Pew Research Center's nationally representative, probability-based American Trends Panel. While trends in candidate standing during the primaries at the aggregate level may have at times appeared linear, the ATP data suggest that voter preferences were quite fluid – with Trump and Clinton both experiencing significant churn in the composition of their supporters. This longitudinal analysis also will examine the demographic groups more likely to support a candidate consistently, and those that wavered in their support throughout the campaign. Additionally, it will analyze differences in the general election preferences of those who expressed more and less consistent support of their party's eventual nominee in the primary season.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Attitudes Toward Democracy and the 2016 Presidential Election

Kirby Goidel *Texas A&M University*

Keith Gaddie *Oklahoma University*

To what extent do attitudes toward democracy shift over the course of an election campaign? In this paper, we, first, investigate differences in citizen attitudes toward democracy prior to the 2016 election, paying particularly attention to differences based on partisanship, ideology, and media exposure. Second, we examine how these attitudes shifted in light of Donald Trump's surprising victory. Did the victory make his supporters more supportive of democratic governance? Did Hillary Clinton's loss influence their attitudes toward democracy? And, what role if any, did the media play in this process? To investigate these questions, we utilize original data from a GfK Knowledge Panel and from the Cooperative Congressional Election Study.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

The Effect of Interview Duration on the Think-Aloud Quality in Children

Mila Sugovic *EurekaFacts, LLC*
Michael Plotkin *EurekaFacts, LLC*
Bohdana Sherehiy *EurekaFacts, LLC*

This study evaluates the effect of interview duration and practice on the quality of verbalization of information in a think-aloud method used during cognitive interviewing. Research suggests that young children typically find the think-aloud method to be difficult (Somerén et al, 1994), however, it is unclear whether the quality of verbal reporting produced by children can vary as a result of practice. Past research shows that there is no difference in the effectiveness of the think-aloud method in children of different ages (Sugovic et al, 2016). However, there is little empirical research on the effect of interview time and practice on the quality of the think-aloud in children of different age groups. This study examines the application of a think-aloud technique in survey-item testing across two different age-groups: grade 4 (age 9-10) and grade 8 (age 14-15). We compared the effectiveness of the think-aloud method between the two grades by evaluating the changes in verbal reports length, response duration, number of issues identified and the quality of student responses over the course of an interview. We found that the effect of time on the quality of the think-aloud varies between the two age groups. More specifically, verbalization of information improves over time in children in grade 4 but not for those in grade 8. This research has important practical implications as it suggests that younger students may need additional think-aloud practice time in order to optimize the quality of the think-aloud for items relevant to testing.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Measuring Patient Satisfaction And Quality Of Care From a Hospitalist Perspective

Veronica B Hoyo *UCSD- CTRI*
Daniel Bouland *UCSD*

In a more competitive, better communicated, globalized world, patients face an unparalleled number of options for their healthcare. With an increasingly diverse offer of services and greater information about diseases and pathologies, patients have developed higher expectations, they have become more inquisitive consumers but, paradoxically, they may not always have the means to gauge the quality of their healthcare alternatives. Similarly, healthcare organizations striving to succeed need to make ends meet by reducing costs, increasing profitability and delivering exceptional medical services. What is needed now is a tool to promote value-based competition on patient-centered results (Porter & Teisberg, 2004). Our planned survey instrument is an attempt to shed light on the measurement of patient experiences by focusing on patient-reported data on satisfaction and quality of care from a Hospitalist perspective. Hospitalists -defined as specialists in inpatient medicine- (Wachter & Goldman, 1996), offer an interesting methodological challenge due to the particular nature of their practice: they focus on patients, and their needs, exclusively during their hospital stays which limits their ability to have a long-lasting relationship with them. The survey instrument we propose addresses several problems that have been identified in previously used satisfaction questionnaires carried out in an academic healthcare system, namely: the indistinguishability of the Hospitalist and the care they provide to the patient; gaps in communication and in matching patients' expectations; the presence of "grading-drift" (hospitalists being blamed for services they do not provide ie., nurses or custodial); and basic accountability problems given length of stay, rotations and shift-work. Our pilot questionnaire will be designed to incorporate questions pertaining to safety, timeliness, efficiency, equity and effectiveness of care delivery (IOM standards). Pretesting and implementation strategies will be discussed too.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

An Investigation of Attitudes Toward Social Media Use in Local Government

Xiaoheng Wang *University of Illinois at Chicago*

Allyson Holbrook *University of Illinois - Chicago*

Mary Feeney *Arizona State University*

Social media technologies (i.e., Facebook, Twitter, YouTube, LinkedIn) have been widely adopted by governments as a way to increase civic engagement, promote governments' openness, and extend government services (Mergel, 2010; Oliveira & Welch, 2013; Yavuz & Welch, 2014). However, government interactions with the public through social media tools has also introduced challenges related to information security and management issues for government privacy policies and requirements (Jaeger et al., 2012). This paper aims to examine government officials' opinions and beliefs about social media use in their departments, and to examine whether they vary as a function of the type and the characteristics of the departments in local governments, as well as the characteristics of the city. We use data from a 2014 national web survey in the U.S. on technology in city government. The survey was administered to municipal officials in five positions — City Manager/City Administrator, Director of Community and/or Economic Department, Finance Director, Director of Parks and Recreation, and Deputy Police Chief — in 500 local governments with population ranging from 25,000 to 250,000. A total of 739 responses were received for a final response rate of 33% (AAPOR Response Rate Calculator V3.1). This survey includes 14 items that use a five-level Likert response scale (strongly disagree to strongly agree) that measure respondents' opinions of social media use and organizational environment (i.e., work routineness and centralization). We will use a few analysis strategies including factor analysis and multilevel modeling to investigate the structure of municipal officials' attitudes toward social media use and whether they vary by different departments or different cities. The results will be important for understanding how social media technologies are perceived and used by government officials and have implications for public outcomes.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

2016 National Survey of Children's Health Incentive Effectiveness

Lauren DiFiglia *U.S. Census Bureau*

Lauren DiFiglia *U.S. Census Bureau*

Jessica Jones *HHS/HRSA/MCHB/OER/DE*

The U.S. Census Bureau administers the National Survey of Children's Health (NSCH) annually on behalf of the Maternal and Child Health Bureau (MCHB), within the Health Resources and Services Administration (HRSA), Department of Health and Human Services (HHS). Data collection strategies were designed to promote survey participation, including the use of incentives. As part of the 2016 production survey, sampled addresses received a \$2 or \$5 cash incentive or were part of the control group that did not receive a cash incentive. Survey research indicates that incentives are a necessary and cost-effective expense for achieving a response rate that minimizes non-response bias. Accordingly, incentives were used in all previous administrations of the NSCH/NS-CSHCN, and tested in the NSCH 2015 Pretest. Results from the 2016 NSCH will be used to evaluate the effectiveness of the incentives by comparing overall response rates and the cost of data collection between each treatment. We will specifically calculate the cost per household, screener return, and topical return for each treatment. We will also compare the treatments by sample characteristics, including households with children versus households without children, to see if they respond differently, in level and timing, to the varying incentive amounts. While data collection for the 2016 NSCH will not be complete until February 2017, preliminary response rates show that there is a statistically significant increase in response rates when respondents were provided a \$2 incentive compared to those in the control group. Although there was a slightly larger increase in response for households mailed a \$5 incentive compared to those mailed a \$2 incentive, this treatment may not be cost effective. Further research will be conducted to compare the response and cost effectiveness of the incentives for specific sample characteristics as data collection operations come to a close.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Motivation to Participate in—and Complete—Election Polls

Erin Pinkus *SurveyMonkey*
Laura Wronski *SurveyMonkey*
Mark Blumenthal *SurveyMonkey*
Jack Chen *SurveyMonkey*

SurveyMonkey's Election Tracking fielded a continuous election poll at an unprecedented scale, interviewing more than 1 million registered voters over the course of the campaign. To do so, we employed a unique method of recruitment: presenting a random sample of respondents to user-generated SurveyMonkey surveys (of which we have more than 3 million respondents each day) with the opportunity to participate in our election poll. In this study, we examine respondents' motivations to respond to our election survey. We know a lot of information about our respondents: who they were planning to vote for, where they live, whether they are registered to vote, and so on. We have less information about our non-respondents; but, one piece of information we have that can be particularly insightful is the type of survey a respondent or non-respondent completed just prior to the prompt to take our elections poll. After all, these poll non-respondents are motivated enough to complete one survey—whether it was sent to them from a friend, coworker, or anyone else—but not a second. We are interested in the potential for differential response and in singling out what—if any—characteristics of these “feeder” surveys might make someone more or less likely to respond to our elections poll. There is a great deal of literature demonstrating that survey respondents are more likely to be actively engaged in society in other ways: through volunteering, church attendance, and even voting. This research will provide evidence for whether there exists some additional factor that distinguishes typical survey respondents from election poll respondents in 2016, and whether that differentiation had an effect on our final results.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Reviewing the Results: Sampling and Survey Design with Administrative Records Supplementation in the 2016 National Survey of Children's Health

Scott Albrecht *U.S. Census Bureau*

Scott Albrecht *U.S. Census Bureau*

Jason Fields *U.S. Census Bureau - Washington, DC*

Reem Ghandour *HHS/HRSA/MCHB/OER/DE*

Jessica Jones *HHS/HRSA/MCHB/OER/DE*

The National Survey of Children's Health is directed by the Maternal and Child Health Bureau (MCHB), within the Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services (HHS), and administered annually by the Census Bureau. The survey provides national and state representative estimates for children aged 0-17 years with and without special health care needs. For 2016, Census developed an integrated sampling strategy that supplemented the Master Address File (MAF) with administrative records. This strategy allowed for two key innovations. First, we integrated indicators from administrative data to assign a child-present flag to addresses within the Master Address File. Second, we utilized paradata to designate geographies as 'internet-restricted', and adapted nonresponse follow-up based on that flag. The child-present flags designated households with at least one child under the age of 18, and was developed from the Social Security Administration's (SSA) Numident and other sources. Internal audit of the flag indicated that children were present at more than 70% of flagged addresses and less than 10% of other addresses. We estimated that a 5:1 oversampling of flagged households over non-flagged households would double the proportion of households with children in the sample over the base proportion in the population. Data collection for the 2016 NSCH is ongoing; however, current results are in line with expectations, as 76% of flagged households have reported children present versus 7% of other households. Second, we developed a flag to identify 'internet-restricted' households using tract-level web response to the American Community Survey. We initially invited all households to respond by web; however, non-respondent households flagged as 'internet-restricted' were diverted to paper mailings earlier than 'internet-enabled' households. Early results indicate the flag is very effective at identifying web-preference but not paper-preference. Additional inputs are required to more efficiently target addresses by survey mode.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

A Bayesian Approach to Remedy the Consequences of Within Subject Correlations in Mode Effect Adjustments

Heather Hisako Kitada *Oregon State University*

Sarah C. Emerson *Oregon State University*

Claudio Fuentes *Oregon State University*

Originally identified in Deming's seminal taxonomy as "factors that affect the usefulness of surveys" in 1944, the changing landscape of survey mode administration is still a dynamic area of research. The topic of mode effect has been well established within sociological and psychological literature; however, a recent surge of large, repeated national surveys transitioning to mixed-mode designs has necessitated the need for statistically supported adjustments. Several ad hoc statistical attempts have been made to adjust for survey mode effects. These adjustment methods are based on the framework of creating counterfactuals. Thus reassigning respondents to one mode in order to establish consistency for ease of comparison through time. Most of these procedures assume no correlation between an individual's hypothetical responses to different modes. However, the implied utilities latent logistic multiple imputation algorithm assumes the strongest possible correlation (proposed by Kolenikov and Kennedy 2014). Hence, there is a gap in the literature regarding intermediate correlation. Although individual effects are inestimable within a single survey study, the definition and treatment of correlation structure affects the conclusions we draw and how accurate they are. The goal of this paper to create a flexible unified correction model that explicitly parameterizes within respondent correlation, i.e., a random person effect. We also approach this problem from a Bayesian perspective by including a prior on the within subject effect (correlation structure). The effect of this parameter will be evaluated through simulations and sensitivity analyses. Finally, this method will be applied to the National Oceanic and Atmospheric Administration's (NOAA) pilot study to explore the transition of the Fishing Effort Survey (FES) from using RDD telephone administration with the Coastal Household Telephone Survey (CHTS) to a mail survey design.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Measuring Financial Literacy in a Large-Scale General Survey

Jonas Frederik Beste *Institute of Employment Research*
Arne Bethmann *DJI*

In a variety of research questions, e. g. the risk of poverty or the relationship between income and living conditions, the ability of a person to use knowledge and skills to effectively manage financial resources is a relevant factor. A person who exhibits a high degree of this ability is likely to have a lower risk of poverty, a more efficient conversion of income into living standard and a higher satisfaction with his income. This ability is described by the concept of financial literacy. Although, it is substantial for many research projects, financial literacy is seldom measured directly in social surveys. Often education is used as a proxy. Previous research has shown that education is a weak proxy for financial literacy. To take the financial literacy of a person into account investigating this kind of research questions, we developed an eight item Likert scale which ran in 2014 on the 8th wave of the Panel study 'Labour Market and Social Security' (PASS). PASS is a longitudinal data set for Germany that focuses on welfare receipt and labour market participation but is also usable to give evidence about a variety of subjects for the German general population. The question battery covers three theoretically dimensions: planning of finances, handling current finances and financial pressure. In this presentation we introduce our measurement instrument for financial literacy. To verify that the same construct is measured across different groups, we perform tests of measurement invariance. Here, measurement invariance is tested in the framework of multi-group confirmatory factor analysis (CFA). We focus on differences between person with low and high income as well as person with low and high education. It could be assumed that those groups differ in the makeup of the underlying construct so the mean-values could not be simply compared to each other.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Importance of Qualitative Assessment to Inform Quantitative Measurement and Guide Interventions During Project Implementation

Ali H Mokdad *University of Washington*
Charbel El Bcheraoui *University of Washington*
Erin Palmisano *University of Washington*
Alex Woldeab *University of Washington*
Emily Dansereau *University of Washington*
Alex Schaefer *University of Washington*
Bernardo Hernandez *University of Washington*

The Salud Mesoamérica Initiative (SMI) is a results-based initiative aimed at reducing inequalities in maternal and child health in rural Central America and Southern Mexico. The initiative includes a baseline assessment to set targets and three follow-up measurements at 18-24 month intervals to assess achievement of targets and release performance-based payments. Key indicators include coverage of contraceptives, coverage and quality of antenatal and postnatal care, in-facility delivery, skilled birth attendance, management of maternal and neonatal complications, complete vaccination coverage for age, and prevalence of anemia in children. In addition to quantitative surveys in each country, we conducted a qualitative assessment in Mexico to complement the findings and inform the Ministry of Health and others on means to improve the intervention. The baseline survey in Mexico was conducted in 5,428 households and 90 facilities. The first follow-up measurement was conducted in 60 facilities. The qualitative evaluation included 120 in-person interviews with key stakeholders and 45 group discussions with community members. Participants included representatives from Federal and State Ministries of Health, SMI Donors, the Inter-American Development Bank, technical assistance partners, health care providers, midwives, and community members. Quantitative surveys showed tremendous improvement from baseline to first follow-up measurements. For example, hospitals with availability of essential inputs for delivery and newborn care increased from 6% to 100%, within two years. Qualitative investigation showed improvements in facility management and protocols, supply logistics, and ministry responsiveness towards facilities, though lack of human resources and social unrest are identified as barriers for improved health services in the region. Our study shows the importance of qualitative data to inform and guide interventions during project implementation. We identified many issues that need to be addressed in the intervention. Moreover, qualitative results informed us about additional questions and areas to be included in the following rounds of quantitative surveys.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Sexual Identity and Sexual Behaviors among High School Youth

Sarah Conklin Virginia Department of Health

Data from the Virginia Youth Survey, a population-based surveillance system that monitors health-related behaviors of youth, were analyzed to examine the relationship between sexual identity and sexual intercourse behaviors. For the first time in Virginia, an urban sub-site opted to include sexual behavior questions as part of the Virginia Youth Survey. High school students ($N = 1,113$; $m = 592$, $f = 516$) from an urban school district in Virginia completed the survey in the fall of 2015. The majority of the sample was African American (79.6%), followed by Hispanic (10.6%), and then White (7.9%). Sexual identity was assessed using the question "Which of the following best describes you?" Answer options were heterosexual, gay or lesbian, bisexual, or not sure. Students predominantly identified as heterosexual (79.7%); however, 15.1% of students self-identified as gay, lesbian, or bisexual. Logistic regression analyses were conducted to examine the relationship between sexual identity, sexual intercourse, and condom use. Sexual intercourse included two different variables – lifetime sexual activity and current sexual activity (one partner in the previous three months). Students who identified as gay, lesbian, or bisexual were 2.1 times (95% CI [1.4, 3.1], $p < .001$) as likely to report ever having sexual intercourse, and 1.7 times (95% CI [1.2, 2.6], $p < .01$) as likely to report current sexual intercourse. Of the students who reported current sexual activity, students who identified as gay, lesbian, or bisexual were .4 times (95% CI [.2, .7], $p < .01$) as likely to report condom use. This is the first time Virginia has been able to examine the relationship between sexual intercourse, sexual identity, and condom use using data from the Virginia Youth Survey. The findings demonstrate the importance of measuring youth sexual behaviors using anonymous, survey-based research. Continued surveillance is needed to better understand the behaviors of a potentially vulnerable population of students.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Practical MRP: Beyond Political Polling Applications

Robert A Petrin *Ipsos Public Affairs*

Alan Roshwalb *Ipsos Public Affairs*

Joe Zappa *Ipsos Public Affairs*

Zachary Lewis *Ipsos Public Affairs*

Multilevel regression with poststratification (MRP) has been shown effective in making statistically valid population inferences with data drawn from nonprobability samples (Gelman, 2007; Ghitza and Gelman, 2013). Multilevel regression with marginal poststratification (MRmP) has emerged as a compelling, scaled-down version of MRP applicable in situations where the volume of population data MRP requires are not available to allow a deep post-stratification on adjustment variables (e.g., Leemann and Wasserfallen, 2014). In spite of the promise of these methods, many of the published examples of MRP and MRmP involve political outcomes for which very strong predictors are available for poststratifying (such as Party ID for use in estimating presidential approval or vote preference). As a result, public opinion researchers have little data on how well these methods work outside comparatively “ideal” situations. The purpose of this paper is to evaluate the performance of MRP and competing methods when applied to non-political measures, as well as in situations where the number of survey outcome measures of interest is large, making item-by-item MRP or other adjustments unreasonable. Performance will be assessed using survey items on topics ranging from health and substance abuse, attitudes toward higher education, altruistic behaviour, and confidence in the US economy. Benchmark response levels will be drawn from the National Survey on Drug Use and Health (NSDUH), Sallie Mae’s How American Pays for College (an annual national telephone survey using RDD among students and parents of college students), the Health Interview National Trends Survey (HINTS), and the General Social Survey (GSS), among others. The methods that will be compared will include MRP, MRmP, and more generalized Bayesian methods for integrating sample information into model-based design adjustments (e.g., Gelman et al., 2013; Pfefferman, 2011; Reilly et al., 2001), as well as propensity score matching (Guo and Fraser, 2015).

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Analysis of Factors Affecting Response to Web Surveys

Lena Le *Social and Economic Sciences Research Center*
Thom Allen *Social and Economic Sciences Research Center*

The Social and Economic Sciences Research Center (SESRC) at Washington State University has been conducting web surveys for the past twenty years using a variety of techniques and procedures in order to meet project specific needs while maximizing response rates. With over five hundred web surveys conducted during this period of time, a rich pool of data exists from which to analyze the factors that have the greatest impact on rates of response. Using a meta-analysis of these surveys, we are examining how the use of incentives impacted rates of response during the past two decades, including the contribution of pre versus post incentives, the use of cash and non-cash incentives and how different monetary amounts had an impact on response. We also analyze the effect of the contact sequence on response, including the number of contacts, the contact modes and combination of modes, the timing of contacts, as well as the types of appeals used to encourage response. In this poster we also look at the web survey as a stand-alone mode versus its share of response within a mixed-mode design, including effect size in regards to its role in simultaneous and sequentially administered mixed-mode surveys. Other factors, such as the survey's sponsor and its geographical relevance to the study population, the population size and type, the time of year the survey was administered, the length of the survey, and the levels of personalization are also examined. Because all of the surveys analyzed in this poster were administered at the same organization and within the same general methodology (TDM), some elements of the survey design remain constant across all studies, reducing the effect of confounding variables and providing a unique perspective on the effects of those elements that actually differed.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Examining Multi-Cultural Differences in Attitudes about Quality of Healthcare Systems

Orin Puniello *Ketchum Global Research & Analytics*
Hillary Schuetz *Ketchum Global Research & Analytics*

A number of organizations, including the World Health Organization and the Common Wealth Fund, provide global rankings of healthcare systems. These rankings are usually based on governmental healthcare performance indicators (e.g per capita GDP healthcare spending,). However, these rankings/studies often do not examine the public's perceptions and lived experiences of the healthcare system within the country. This paper seeks to fill this gap in the literature by exploring public perceptions and experiences of healthcare globally, and identify cultural difference that exist in these attitudes. A 2016 study of global healthcare by a major health tech company explored attitudes of patients and healthcare professionals in 13 countries. This survey provides a unique opportunity to not only understand global attitudes about healthcare, but also measure cultural differences between countries. Furthermore, this study examined not only the perceptions of the patient population in each country, but also that of healthcare professionals in each country. Therefore, this study allows for the examination of in-country and between-country differences of attitudes of consumers (patients) and suppliers (healthcare professionals). In total the study provides a robust sample size with 25,000 responses for patients (on average, 2,000 completes per country) and 2,700 response from healthcare professionals (on average, 200 completes per country). Initial results indicate that there are statistically significant cultural difference in attitudes of healthcare systems including key attributes of quality, trust, and patient experience. Further, these data indicate a tension between perceptions of patients and healthcare providers. This paper use multivariate regressions to further understand how cultural differences affect perceptions and experiences of healthcare. The results of this paper will give governments, health advocates, and researchers important context when developing international healthcare policy. The results will be particularly useful in developing multi-country healthcare communications strategies.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Studying Trump-like voters in Mexico

Marco Morales *Instituto Tecnológico Autónomo de México*

Rene Bautista-Martinez *NORC at the Univ of Chicago*

Javier Marquez *Buendía y Laredo*

Jose Merino *Data4*

The recent election of Donald Trump underscored the impact that a coalition of voters that may rally to support a platform based on fear, discrimination, and xenophobia may have. This may also be particularly true for candidates portraying themselves as outsiders to the prevailing political elites in a country. Trump is not an isolated case as many candidates with similar platforms have been recently elected in Europe and Asia, which may preface many more to come. Our concern is with starting to isolate characteristics that voters who support these platforms share cross-nationally. We start exploring this issue with a nationally representative survey collected in Mexico right after the election of Donald Trump. With this dataset, we seek to isolate a potential latent dimension that may identify potential supporters of these issues in Mexico, and scope out the characteristics of these voters. This becomes particularly relevant as a presidential election will take place in 2018, where the temptation to bring platforms with these issues may be big for external candidates.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Use of an Adaptive Research Design to Maximize Response Rate and Data Quality in a Small Population Establishment Survey

Sadie J Bennett *ARDX*

Survey researchers are increasingly employing adaptive research designs to mitigate consistently declining response rates and increases in nonresponse bias. Although there are multiple approaches for incorporation of an adaptive design, much of the current research focuses on responsive design—procedures used to strategically adjust data collection methodology based on the results of a previous data collection phase. This empirical study examines the effect of an adaptive research approach (responsive design) on reach rates, response rates, and ultimately nonresponse bias for a survey conducted on behalf of the Community Development Financial Institutions Fund (CDFI Fund). The CDFI Fund was established within the U.S. Department of the Treasury to promote community development in low-income communities, and encourage revitalization in economically distressed census tracts. The survey was conducted as one component of a baseline analysis of the Bank Enterprise Award (BEA) Program—one of a wide range of initiatives that the CDFI Fund sponsors to promote access to capital and stimulate local economic growth in distressed communities. The survey included representatives from the population of Federal Deposit Insurance Corporation (FDIC)-insured institutions that applied for BEA Program awards during award years 2012-2014. This paper will discuss the utilization of a simultaneous mixed-mode (email/telephone) data collection approach during the follow-up phase of the survey, and the effect of this data collection methodology on overall survey performance, and survey performance by stakeholder category (bank size and BEA awardee vs. non-awardee status). The paper will also examine the effect of the data collection approach on survey performance among institutions providing indirect CDFI-related activities versus those providing direct activities (Distressed Community Financing Activities and Service Activities).

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Modeling Survey Completion Rates and Completion Time using Large-Scale Data

Timothy Rubin *SurveyMonkey*

In this study, we develop a set of machine learning models for predicting two key survey attributes: survey completion rate (defined as the proportion of all respondents that open a survey who end up submitting a completed survey) and completion time (defined as the elapsed time between respondents first opening a survey and respondents submitting a completed survey). These models were trained using a large dataset consisting of tens of thousands of user-created SurveyMonkey surveys, each of which had at least 50 unique respondents. We consider a wide range of survey features to include in our prediction models, including counts of distinct question types (e.g., the total number of multi-choice or matrix questions in the survey) as well as word counts of various components of the survey (e.g., the total number of words in all question-headers in the survey). In the analysis of our models, we consider how accurately each individual feature can predict both outcome variables, as well as how accurately models trained on multiple features can predict both outcome variables. We discuss the implications of these models for the field of survey research, and how these models can be used to guide the creation of higher quality surveys.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Know your Audience: Improving User Engagement measurement in Online Communities of Practice

Maria Payri *American Institutes for Research*

Lauren C'deBaca *American Institutes for Research*

Caitlin Deal *American Institutes for Research*

Communities of practice enable practitioners within a specific professional field to gain access to information and resources that might not be available locally. With a growth in online communities of practice, it is increasingly important for researchers to efficiently and correctly measure user engagement in order to provide members with relevant information, resources, and networking tools. This study examines data from an online community of practice for K-12 teachers. The community has over 4,000 members from all around the US who participate in the online platform in different ways (e.g., being part of affinity-based groups, posting resources and comments). This study aims to go beyond traditional measurement of online activity such as sessions and page views on Google Analytics. First, we will look at trends in joining the community and the relationship with different events during the year, as well as the geographical distribution of the community members by mapping their locations using ArcGIS. Specifically, we will explore user engagement by investigating activities on the community of practice site, such as number of posts by users. We will look at surges in membership after workshops to see if they lead to increased interaction with the community. We will examine whether the geographic distribution of these members is homogeneous or heterogeneous across the United States, and investigate the reasons behind any spatial patterns we find. The findings from this research will inform improvements on how to effectively measure the way we engage individuals in online communities, as well as share an understanding in how these communities develop and interact over time.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

“Mental retardation” vs. “Intellectual disability”: A Natural Experiment

Marina Stavrakantonaki *Univ of Illinois at Chicago*

Timothy Johnson *University of Illinois - Chicago*

We employ individual- and aggregate-level data from 19 years of the National Health Interview Survey (NHIS) to investigate the effects of a legislatively mandated change in the use of medical terminology on self-reports of an important mental health measure. This case focuses specifically on a change from use of the term “mental retardation” to “intellectual disability” that became effective as of 2011 after President Obama signed legislation in late 2010 requiring the federal government to replace the term mental retardation with intellectual disability in many areas of government. This measure, also known as Rosa’s Law, erases the terms mental retardation and mentally retarded from federal health, education, and labor policy. “Intellectual disability” or “individual with an intellectual disability” were introduced in place of these terms, which were considered to be hurtful language. Our analyses focus on changes in national prevalence estimates of this construct during the period 2011-2015, compared to 1997-2010, for the population as a whole, and also for demographic subgroups, using a pooled dataset and multivariate models designed to isolate potential trend changes that can be attributed to the introduction of the revised question wording. These data are additionally aggregated into quarterly prevalence estimates that provide 76 data points for use in a set of interrupted time series analyses to further evaluate the impact of the 2010 legislation. At both levels of analyses, we also examine estimates of survey reporting of other health conditions that are employed as controls for further isolating the effects of this natural experiment.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Managing Design Effect in Non-Probability Samples

Austin Albino Nielsen - *San Francisco, CA*
Lanie Anton Nielsen - *San Francisco, CA*
Alex Fragapane Nielsen - *San Francisco, CA*
Sharon Turlington Nielsen - *San Francisco, CA*

Design Effect (and its impact on effective sample size) due to weighting or sample clustering (Kish, 1965) is a known issue in methodological circles, but often not given due consideration in research design and power analysis prior to data collection. Methods for considering, and mitigating, threats to effective sample size are discussed. The authors describe the starkly non-linear relationship between the degree of bias vs. weighting targets and Design Effect. Subsequently discussed is the tradeoff inherent in weighting dimensions that reduce bias, but also reduce statistical power (and therefore precision). Authors suggest that researchers carefully balance reduction in bias with loss of power when designing a weighting scheme. Analyses and simulations are conducted using a current market research sample (N~100K), examining the impact of various weighting factors on the validity and reliability of survey research results.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Online Non-probability vs RDD in Statewide Public Interest Surveys

Krista Jenkins *Fairleigh Dickinson Univ*
Peter Woolley *Fairleigh Dickinson University*
Dan Cassino *Fairleigh Dickinson Univ*

The Fairleigh Dickinson University Poll, PublicMind, was created in 2001. Its work in measuring attitudes and behaviors largely at the state level (New Jersey) has relied exclusively on the “gold standard” in survey methodology – probability samples utilizing RDD. Ongoing challenges to RDD surveys abound (cell phones, TCPA, declining response errors and non-response bias, etc.), necessitating an exploration of other measurement strategies. This poster will add to the growing literature comparing data gathered through RDD telephone surveys and opt-in, internet panels. We will simultaneously conduct two surveys – one utilizing RDD (cell and landline) and the other drawn from online interviews using a non-probability panel (Qualtrics). The same questions that we routinely ask on our statewide surveys (gubernatorial job performance, direction of the state, etc.) will be asked of online respondents, and, to the extent possible, the online respondents will be identified to match relevant known demographics of the New Jersey population. Questions will also be drawn from those that measure known behaviors as measured on government surveys. The study provides further exploration of differences in findings across disparate modes of data collection, and the extent to which the more cost-efficient online strategy provides an alternative to the RDD “gold standard.”

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Are You Still Online? Measuring Internet Access from Home for School-Age Children

Angelina KewalRamani *American Institutes for Research*

In our advancing technological society, school-age children increasingly need access to computers and the Internet at home in order to progress through their education. In 2015, approximately 81 percent of children under age 18 in the United States lived in a house with Internet access (ACS 2015). However, prior research has shown that disparities in home Internet access exist by poverty, geographic region, race/ethnicity, and parent's educational attainment. Therefore, it is important to be able to examine differences in Internet access for small subgroups of school-age children. Different data sources are available for measuring Internet access from home and each offers advantages and disadvantages. Since 1997, the U.S. Census Bureau has reported on Internet use from home using the Current Population Survey (CPS) supplements. This paper will study the recent period of 2010 to 2015 when questions on Internet access were more consistent. Based on the July and October CPS supplements, Internet access from home for school-age children fluctuated between 2010 and 2015 with no clear pattern of increase. This paper will examine these puzzling findings and investigate whether context effects may be a factor. The July CPS supplement included a detailed battery of questions on household and individual Internet use, while the October CPS supplement included a short subset of questions on these topics. In addition, the Census Bureau has reported on Internet access from home starting in 2013 using the American Community Survey (ACS). The CPS and ACS differ in several aspects, including sample size, mode of administration, questionnaire design, and reference period. This paper will investigate fluctuations in CPS estimates of Internet access for school-age children between 2010 and 2015 and also present recent ACS estimates on this topic by different demographic characteristics. The analyses will help researchers determine which data source best fits their needs.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Exploring Political Epistemology Among Local Republican Committee Members

Heather Knappen

The long-accepted methods for political polling, particularly by telephone, are increasingly up for challenge. The election of Donald Trump to the American Presidency is just a recent example in which political pollsters have had to account for (and defend) questions from the public and media related to inadequate sampling coverage and declining participation rates. Another challenge heard less frequently is related to the quality of survey responses. Even if survey participants represent the “right” demographic and 100% of them participate in a poll, this is no guarantee that they possess adequate knowledge of the subject matter and are able offer reliable opinions. This poster explores a unique form of registration-based sampling by focusing on town-level Republican committee members within Monroe County, NY. Town-level committee members are registered in the Republican Party and elected to their committee position. In this capacity, they represent registered Republican voters in specific Election Districts (ED’s) throughout the county. As “political insiders,” it was theorized that these individuals would be more likely to participate in a political survey and provide expert knowledge and insight when compared to a less-engaged voter. In March, 2016, Republican committee members from two townships were invited to participate in an on-line survey. The survey asked each member who they planned to support for President in the New York State Republican Primary Election and which candidate they expected to win (regardless of who they supported). More than half of the committee members completed the survey and their responses reflected the true election outcome within Monroe County. As the AAPOR community continues to embrace change and diversity in public opinion research, this poster presents an alternative approach to sampling for consideration.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Let's Recruit an Internet Panel: Multiple Tries at Methods

Jennifer Hunsecker *Nielsen - Oldsmar, FL*
Lauren Walton *Nielsen*
Kay Ricci *Nielsen*
Amanda Tscheiner *Nielsen*

Recruiting and retaining an Internet panel can be a challenging endeavor. This study will describe four recruitment studies designed to increase Nielsen's computer and mobile device panels. These studies took place over a five month period and illustrate the variety of methods that online panels managers use to grow their membership. Google AdWords is a service that allows clients to write an ad that will appear when certain search terms are typed into Google. The primary purpose of the AdWords project was to recruit Spanish-speaking Latinos to the Nielsen online panel. This recruitment ran from the end of June 2016 until the beginning of September 2016. The ads received approximately 804,000 impressions, 12 clicks, and cost approximately \$15,000. The total number of panelists recruited was 73. While cross-recruitment from panels is often used by survey researchers, the results in the computer panel have been mixed. One outreach invited members of an internal survey panel to join the Nielsen online panel. This effort proved to be unsuccessful, with nearly 80,000 email invitations sent for 180 new panelists. On the other hand, the digital recruitment team has had luck with recruiting internally from the computer panel to the mobile measurement panel. Members of the computer panel are sent an email extending an invitation to join the mobile phone panel. This low-cost, low-time investment effort has added an additional 75-100 panelists a month. A markedly different effort involved surveying former computer panelists. A large online sample provider recontacted members exposed to the Nielsen panel offer who did not sign up. Unfortunately there were few complete survey responses received (150 completed of 1520 invitations). Responding panelists indicated confusion over the registration process and a desire for more incentives. Further information on respondents, results, and lessons learned will be provided.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Small Group Recruitment Strategies with a H2R Population

David P Getman *Decision Information Resources, Inc.*

Sylvia Epps *Decision Information Resources, Inc.*

Ron Bass *MDRC*

Research has shown that non-probability recruitment strategies asking hard-to-reach participants to refer people they know, such as snowball or respondent-driven sampling, are both cost efficient and effective. (Patrick et al, 1998; Heckathorn, 2011). Similarly, nonprobability web-based surveys have been shown to benefit from respondents forwarding study information to people they know, known as the “Pass-Along Effect” (Norman and Russell, 2006). Less is known, however, about how recruiting hard-to-reach sample in small groups of respondents known to one another beforehand affects response rates for longitudinal probability studies, which is the focus of the proposed study. Specifically, using two waves of data from a three-wave study of a microcredit/microfinance group-lending program, this proposed study will examine intra-group reporting among predominantly Spanish speaking female loan applicants who were asked at 6 month and 18 month follow ups to confirm or update the contact information and to rate their closeness to other sample members they were grouped with at recruitment. We will address the following research questions: • Do respondent indications of group closeness correlate with the group’s tendency to complete as a group? • Did early completes within groups influence other group members’ responsiveness? • Does group responsiveness vary by study group (treatment/control)? The results will be inform similar large-scale evaluations on an understudied topic and population in the U.S. Additional analyses will be conducted to determine whether these early efforts impacted responsiveness over the duration of the 18-month and upcoming 36-month follow-ups.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Immigration, Integration, Religious Tolerance and Identity: Germany in 2008 and 2016

Sofia Pinero Kluch *Gallup*

Kenneth Kluch *Gallup*

Alan Vaux *Southern Illinois University*

With populations around the world encouraging the closing of national borders (such as BREXIT), heightened concerns regarding immigration trends generally; and specifically in the U.S. with fears of potential terrorists using refugee status to infiltrate immigration policies; Germany made headlines in 2016 for taking in hundreds of thousands of Syrian refugees. In 2008 the Gallup World Poll conducted a nationwide CATI study in Germany of both the national population (N=1000 sampled via RDD) and German Muslims (N=500 sampled via lists) covering topics of identity, integration, religious tolerance, and wellbeing. The 2016 fielding of the World Poll in Germany again included these topics along with questions about both domestic and international terrorism within Germany. The current paper explores how attitudes towards individuals of other religious faiths have changes during the eight year period. Views of immigration, civic engagement, and identity – with community, country, ethnic background, religion, and Europe - are all explored. German populations will be compared based on age cohorts, gender, and employment. Further segmentation analyses will identify the populations more positive and favorable to both immigrants coming to Germany and also to those practicing different faiths to determine how or if these segments have shifted over the past eight years. Without a 2016 German Muslim comparison group, the general population sentiment will be used to summarize the current climate and potential shifts in views towards Muslims generally, and immigrants specifically while the 2008 German Muslim data will be used to better describe the state of the Muslim population at the time of the initial data collection.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

10 Years of Stress in America: Keeping Trended Data Relevant

Aimee Vella Ripley *Nielsen Consumer Insights*

Sophie Bethune *American Psychological Association*

Allyssa Birth *Nielsen Consumer Insights*

Stress in America, an annual survey conducted by Harris Poll (Nielsen) on behalf of the American Psychological Association, has just reached its 10 year milestone. In examining the issue of stress in this country, over the past 10 years we have explored public opinion and research on racial, ethnic, religious and sexual orientation issues. At its core, our survey touches on the topics of overall health, attitudes towards stress, assessment of personal stress, the impact of stress, stress management, and the availability of emotional support. While observing the changes in these trends is the crux of our study, every year we are faced with the challenge of balancing trended research with fresh and newsworthy public opinion. The world today is very different than it was 10 years ago and solely focusing on these measures year over year would result in a study that not only becomes repetitive, but runs the risk of becoming irrelevant. Our solution has been to keep our core measures of stress and other health-related issues consistent year-over-year while selectively adding in current events. Over the years this has included examinations of healthcare reform, discrimination, technology, and the 2016 presidential election. The results speak for themselves as our findings related to the presidential election received unprecedented news coverage, resulting in over 3,000 news stories on the topic. This approach has enabled us to continue to contribute to relevant and timely public opinion research without sacrificing trends or becoming stale.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Determinants of Consent and Response Rates for Studies of Cancer Patients: Evidence from Multiple Studies Initiated from a Statewide Cancer Registry

Morgan Millar *University of Utah*
Sandra Edwards *University of Utah*
Carol Sweeney *University of Utah*

Epidemiologists and clinical researchers are increasingly interested in obtaining patient-reported outcomes via surveys of individuals diagnosed with cancer. A potentially promising method of identifying and recruiting eligible cancer survivors for participation in surveys of health and quality of life post-cancer care is through cancer registries, which document patient information for each case of all reportable cancers diagnosed within their registry area. This project reports results of an extensive analysis of consent and survey response rates from 10 years of studies for which the Utah Cancer Registry contacted cancer patients to obtain consent for such studies. We assess determinants of patient consent and response to surveys, including patient demographics, time since diagnosis, cancer site and stage, and study implementation procedures and protocol characteristics. Preliminary analyses using a subset of available data indicate that rates of consent are lower among older patients, and also vary by type of cancer diagnosis. Additionally, we find an inverse relationship between time since diagnosis (in years) and likelihood of consent. Further analyses of determinants of consent outcomes will provide useful insights into barriers to obtaining survey responses from cancer survivors and methods that work well for implementing such surveys, while also adding to the growing methodological literature assessing determinants of survey consent and response rates.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Acquiescence Bias in Yes-No Grids? The Survey Says... No.

Randall K. Thomas *GfK*

Frances M. Barlas *GfK*

Nicole R. Buttermore *GfK*

Jolene D. Smyth *University of Nebraska - Lincoln*

Yes-No Grids (YNG) have been found to yield higher endorsement rates than a Multiple Response Format (MRF - i.e., select all). A number of studies have indicated that higher endorsement for the YNG may be due to increased consideration of each element (the salience hypothesis). Alternatively, Callegaro et al. (2015) proposed that the higher endorsement in the YNG is most likely explained by acquiescence bias, which is the tendency for people to be agreeable and thus more likely to endorse 'agree' in an agree-disagree response format or select 'yes' in a YNG. In an initial test of these alternatives, we found strong support for the salience hypothesis and little support for the acquiescence bias. In the current experiment we expanded from the more narrowly focused YNG to explore the effects for dichotomous grids (DG), which include response options beyond 'Yes' and 'No'. Over 1,500 respondents from GfK's probability-based KnowledgePanel® completed an online survey in which they were assigned to evaluate a series of behaviors with either an MRF (asking 'Did you do each of the following?') or a DG, which asked either 'Does it describe what you did?' (with either 'Yes-No' or 'Describes-Does not describe' responses) or 'Did you do each of the following?' (with either 'Yes-No' or 'Describes what I did-Did not describe what I did' as responses). While all DGs had higher rates of endorsement than the MRF, we found no evidence that a 'Yes-No' format led to increased endorsement over and above the 'describes' response alternative. With our validity criterion (number of days in past 30 they did each behavior), we found slightly higher validity coefficients for the DG items than the MRF. Taken in total, the salience hypothesis was a better explanation for the higher levels of endorsement in DGs and YNGs compared to the MRF.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

New democratic and methodological changes in Latin America: the case of 'anti-systemic vote' in Mexico

Diana Paola Penagos Vasquez *Parametria S.A. de C.V.*

Francisco Abundis Luna *Parametria S.A. de C.V.*

Jose Alberto Vera Mendoza *Parametria S.A. de C.V.*

The most recent elections in England, Greece, Scotland and Brazil have questioned the scope of public opinion study. Mexico is no stranger to this demand. Research has proven the profound crisis of confidence that democracy lives in Latin America, and this is even more pressing in the Mexican case. Particularly, the last elections for federal and local officials unveiled new challenges for both democracy and for the guild pollster in Mexico. People's distrust in the authorities and quality of the elections have set a more critical electorate and more willing to substantial changes. Using data from pre-election polls and exit polls of the most recent elections, this research analyzes and models 3 of the most controversial elections for its results. These scenarios are a small sample of some electorate changes comparable to those observed in other countries. The data reveals features of a more and/or better informed citizenry, tired of the traditional political parties, and much less democratic. The phenomenon of an 'anti-systemic vote' is the scope of the elections on June 7 and, with it, the evidence of an 'inconsistent vote' shows the transformation of the electorate explained by theories such as the 'spiral of silence'. In this new political scenario, the methodological research on the acquisition of voting intention should be a continuous practice in terms of changes in voting behavior.

May 20th, 2017
12:45 PM - 1:45 PM
Concurrent Session P

Poster Session #3

Mixed sampling with non-probability sample (Facebook advertisement) and probability sample (address-based sampling)—does it work?

Chanyoung Lee *University of South Florida*

Communication technologies have lowered the cost of communication, but they have induced the frequency of surveys and resulted in significant survey fatigue among potential respondents. Web-based surveys have become a popular mode, but they generally have a low response rate. To address this issue, there is a growing interest in adopting mixed sampling methods, including non-probability and probability sampling. In this study, we used a web-based survey with a mixed-sampling approach to conduct the 2016 Florida Motorcyclist Survey. The survey website was developed and its link distributed through the following methods: • Address-Based Sampling (ABS) – Postcards with an invitation to complete the web-based survey were sent to a random sample of endorsed motorcyclists. In total, 30,000 addresses were sampled from the Florida Department of Highway Safety and Motor Vehicle (DHSMV) database of endorsed motorcyclists. Each sampled address was contacted once by postcard. • Facebook – A paid advertisement that targeted motorcycle riders ages 18 and older was placed on Facebook for 10 days linking to the web survey. In total, 2,670 survey responses were collected, and 56 records in which respondents answered fewer than 5% of questions (early breakoff) were removed. The Facebook approach yielded a higher early breakoff rate (3%) compare to the ABS approach (1%). The mixed sampling approach was used to improve external validity; however, the attempt was limited to improve the sample size to the extent that the collected sample from Facebook (non-random) was a sufficiently large proportion of the overall population. In addition, this approach limited the internal validity of the findings. The study measured and evaluated the outcome of two different sampling approaches, including cost and demographic information compared to population.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Evaluating Confidentiality Pledges in an Era of Enhanced Cybersecurity (Panel)

Evaluating Confidentiality Pledges in an Era of Enhanced Cybersecurity

Cleo Redline *National Center for Education Statistics*
Casey Eggleston, Erica Olmsted Hawala, Jennifer Edgar
Jacob Bournazian *U.S. Energy Information Administration*
Jennifer Edgar *U.S. Bureau of Labor Statistics*
Casey Eggleston *U.S. Census Bureau*
Erica Olmsted Hawala *U.S. Census Bureau*
Jennifer Hunter Childs *U.S. Census Bureau*
Robin Kaplan *U.S. Bureau of Labor Statistics*
Rebecca L. Morrison *National Center for Science and Engineering Statistics*
Heather Ridolfo *National Agricultural Statistics Service*
Stephanie Willson *National Center for Health Statistics*

Protecting the confidentiality of identifiable information reported in surveys is a fundamental responsibility of survey organizations. Consequently, respondent perceptions on privacy and confidentiality have long been an area of survey research. Under the Confidential Information Protection and Statistical Efficiency Act (CIPSEA), most federal statistical agencies pledge that only sworn employees and agents will have access to identifiable information and that this information will be used for statistical purposes only. However, the Cybersecurity Enhancement Act of 2015 provides the Department of Homeland Security with a mandate to monitor any information stored on or transiting to or from federal statistical information systems for any lawful government purpose. This new mandate conflicts with the pledge from federal statistical agencies under CIPSEA. Given this change in potential access to survey data, it is important to evaluate respondents' reactions to any proposed revisions to confidentiality pledges. This panel brings together work by researchers at federal statistical agencies, from a diverse range of subject matters, who evaluated respondents' reactions to revised pledges using multiple methods: analysis of paradata, cognitive interviews in household and in establishment surveys, and web survey results. There is heightened awareness among evaluators that using multiple methods with different subpopulations often results in a stronger, more complete evaluation than using one method and subpopulation. Such an approach can compensate for the weakness of any individual method and allow for greater representation. The panel will discuss the results from each of the individual methods, followed by a paper that will synthesize and summarize the results using the multi-method approach. The papers in this session represent an array of evaluation methods that are available to researchers when circumstances change abruptly and decisions need to be made quickly.

Do They Read It? Using Paradata to Evaluate the Extent to Which Respondents Attend to Confidentiality Pledge Language

Casey Eggleston, Erica Olmsted Hawala, Jennifer Edgar

As part of an interagency effort to evaluate the effect of potential changes to confidentiality pledge language used by agencies of the Federal Statistical System, the U.S. Bureau of Labor Statistics and the U.S. Census Bureau collected indirect evidence about respondent reactions to variations of confidentiality pledge language incorporating new information about cybersecurity monitoring. In

addition to the explicit measures included in cognitive testing and a web survey, paradata (specifically, eye movements and response times) was also collected to capture responses that may have been subconscious or that respondents did not self-report. Overall, both eye tracking and response time data indicated that, while participants did read the confidentiality language, they did not spend a significant amount of time on it and did not especially seem to fixate on the new versions of the pledge that contained information about cybersecurity monitoring. This presentation summarizes the detailed findings and discusses implications for communicating with respondents.

Do People Understand It? Cognitive Interviewing Assessment of Confidentiality Pledges for Household Surveys

Stephanie Willson, Casey Eggleston, Jennifer Hunter Childs

The next two papers describe how respondents interpret and understand different versions of confidentiality pledge language. This paper focuses on household survey respondents, while the next emphasizes establishment surveys. The confidentiality language tested incorporated requirements of the Cybersecurity Enhancement Act of 2015 into the standard CIPSEA language that is already common among Federal household surveys. Cognitive interviewing methodology was used to evaluate the new pledge. Fifty face-to-face, in-depth interviews were conducted; 30 were conducted by the U.S. Census Bureau and 20 by the National Center for Health Statistics. Interviews were designed to explore respondents' interpretation of and reaction to the new pledge language. Because the new cybersecurity language introduces the idea that government data will be more closely monitored than in the past, concerns existed that respondents may react negatively to the new pledge. However, while some did express cautious attitudes about what the new language meant, the preeminent finding was that respondents had generally neutral or even positive interpretations of and reactions to the new pledge language. Some believed that a high level of scrutiny of Federal data systems by the government already existed. Others believed increased monitoring was necessary and important as cyber threats have become more common. Additionally, various versions of the pledge were tested by each agency. It was found that respondent attitudes were influenced by specific wording. For example, versions of the pledge that used the word "protect" invoked mostly positive reactions while versions that used the word "monitor" tended to invoke more negative reactions. This and other specific findings will be reviewed in more detail.

Do Establishments Understand It? Cognitive Interviewing Assessment of Confidentiality Pledges for Establishment Surveys

Cleo Redline, Jake Bournazian, Jennifer Edgar, Heather Ridolfo

This paper discusses cognitive research that explored how establishment respondents understand and react to proposed revisions to agency confidentiality pledges. Using similar protocols, four government agencies (U.S. Bureau of Labor Statistics, National Agricultural Statistics Service, National Center for Education Statistics and the U.S. Energy Information Administration) conducted 102 cognitive interviews with participants from different types of establishments, including private and public companies, farms, and schools. Participants were probed to determine whether they understood the intended meaning of the new language so that agencies could assess the potential impact on their future data collections. Participants generally understood the proposed language. The research identified differences in interpretations of and reactions to the proposed text by establishment and respondent types. For example, farmers were more likely to be concerned about the new language, while energy companies preferred wording that specifically mentioned DHS. This paper will cover the testing approaches used,

summarize the results and detail the advantage of conducting cognitive testing with unique respondent groups.

Does It Matter? Impact of Confidentiality Pledges on Web Survey Response

Jennifer Edgar, Robin Kaplan, Casey Eggleston

As part of interagency research on the potential impact of changes to confidentiality pledges, two online surveys were conducted. The U.S. Bureau of Labor Statistics (BLS) sent a survey to a sample of 864 participants from Amazon Mechanical Turk, who were randomly assigned to one of two versions of the confidentiality pledge. The U.S. Census Bureau sent a web survey invitation to a national probability sample. Upon starting the survey, participants were randomly assigned to one of five versions of the confidentiality language. For both surveys, total response rates and item response rates were compared by group, as were responses to questions evaluating recall and comprehension of the pledge language. Overall, there were no significant differences by pledge group, participants in both groups completed the survey at equal rates. Additionally, BLS found no demographic differences. Evidence from both surveys suggest that participants generally understood the meaning of the pledge wordings, but either did not understand or remember certain features of some versions. This presentation will summarize the results from both surveys and discuss overall conclusions and implications.

One Size Fits Most? Lessons Learned by Using Multiple Methods to Study Confidentiality Pledges

Heather Ridolfo, Rebecca L. Morrison

In 2016, a workgroup consisting of representatives from ten statistical agencies was established to study respondents' expectations and reactions to changes in the federal confidentiality pledges. The expected outcome of this research was to develop a universal confidentiality pledge that would be used by all statistical agencies. This paper will describe the differences among the various studies (e.g, respondent type, methodology, and pledge wording), then highlight common findings. First, the analysis of paradata and web survey results showed that, although respondents read the confidentiality pledges, they did not spend a lot of time doing so, and often could not recall information provided. Second, the cognitive interviewing assessments showed that participants understood that there would be additional cybersecurity monitoring, but there was inconsistency in the interpretations of who would be doing the monitoring and what would be done with their information. Finally, initial reactions to changes in the confidentiality pledge were neutral or positive, but those reactions shifted upon further reflection. Reactions to pledge changes also varied across survey populations, with the agricultural population having the strongest negative reaction. Ultimately, the panel concluded that a one-size-fits-all pledge was not possible. This presentation will highlight the careful balance that must be struck between a) transparency and clarity and b) respondents' expectations for confidentiality.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Experimentation for Developing Evidence-Based Guidelines for Mobile Survey Instrument Design (Panel)

Experimentation for developing evidence-based guidelines for mobile survey instrument design

Lin Wang *U.S. Census Bureau*

Erica Olmsted-Hawala

Elizabeth Nichols *US Census Bureau*

Erica Olmsted-Hawala *U.S. Census Bureau*

Christopher Antoun *U.S. Census Bureau*

Brian Falcone *U.S. Census Bureau, George Mason University*

Ivonne Figueroa *U.S. Census Bureau, George Mason University*

With the growing use of smartphones and tablets, many surveys can now be administered through mobile devices. A key concern with mobile survey instruments is the optimal design of its user interface to maximize response quality. If not designed appropriately, the user interface of a mobile survey instrument can contribute to measurement error by influencing respondents' perceptions and actions. Optimizing the mobile survey instrument's user interface is thus crucial to reducing measurement error. However, there is a lack of evidence-based user interface design specifications for mobile survey instruments. Inadequate interface design could result in response error, prolonged time in completing a survey, and breakoffs. Further, it could be costly to re-design and re-develop the user interface. To address this important issue, the U.S. Census Bureau is developing evidence-based standards and guidelines for user interface design of mobile survey instruments through empirical research. In this panel, we will first describe the objectives, scope, and strategies of the project. We then report on five experiments that were designed to address practical issues observed in usability testing. The first two experiments investigated how to present information to respondents on a mobile device, while the later three experiments were on how the respondents make responses or enter data. Results from the experiments will become the evidence for the guidelines that are to be developed.

Optimal label location for mobile survey response fields

Erica Olmsted-Hawala

A label of a response field refers to a text that describes the type of data to be entered into the response field, e.g., the word "Age" placed next to the response field in which the respondent's age will be entered. On the mobile device, the label has been placed in various locations including in or outside of the response field. The purpose of the present study is to identify optimal locations at which labels are placed for different data types, so that the respondent can enter the data on the mobile device effectively, efficiently, and with satisfaction. Five possible locations are assessed: above the response field and left-aligned, within the box with dynamic movement above the box when the participant begins typing, left of the response field and left-aligned, left of the response field and right-aligned, and right of the response field. A between-subject design is used in this study. Each study participant will be assigned to one condition and all will perform the same set of data entry tasks on a mobile survey application. Two types of data will be investigated: familiar information, e.g., name, age, date of birth, and unfamiliar information, e.g., the number of hours watching television, cost of electricity, etc. Participants' task performance will be measured with the following performance indicators: task

completion time, errors made during data entry, self-report on the experience of ease of use, and self-report on preference of the label location. We hypothesize that the label located directly above the response field will be most optimal.

Does typographic cueing improve the processing of information from survey questions on a mobile device?

Brian Falcone

As more people use mobile devices to complete Web surveys, it is critical to verify that current Web survey design practices carry over to mobile Web. One set of design practices involves typographic cueing, i.e., using different font styles and sizes to distinguish between the different parts of a question. A common practice is to bold the question stem to draw attention to it and help ensure that it is read first. Another common practice is to italicize instructions to facilitate filtering of information and allow them to be recognized quickly as instructional text. However, typographic cueing has not been tested in mobile Web surveys that typically present one question per page due to the smaller size of mobile device screens. It is possible that the reduction in the amount of information that can be presented at once might reduce or eliminate any benefits of this practice. This talk will present findings from a laboratory experiment ($n=30$) varying two factors: (a) typography of question stems (bolded or not) and (b) typography of instructions (italicized or not). The study questionnaire contains seven questions adapted from national surveys and will be implemented using a native smartphone app. To assess respondent performance with each of the four designs, we compare completion times, response accuracy when answering questions based on vignettes, and post-survey recall of the keywords in the questions. We expect that bolding the question stems will not affect completion time or accuracy because only one question is presented at a time. By contrast, we expect that italicizing the instructions may still assist in filtering which could lead to shorter response times but possibly worse accuracy and recall by providing a cue to respondents that this text may be scanned or skipped rather than carefully read.

Designing response options for touch in mobile Web surveys

Christopher Antoun

As more people use smartphones of small narrow touchscreens to complete Web surveys, researchers have to find ways to format response options in mobile Web questionnaires so that they are easy to touch. Researchers are especially concerned about the errors introduced by radio buttons (for “choose-one” questions) or check boxes (for “choose-all-that-apply” questions) because these icons present a small target area for tapping. Although typically the response option text associated with an icon is a tappable area, some respondents may not know that. One alternative approach is to embed each radio button or check box into a wide button. Another alternative approach is to do away with the icons altogether and present each response option as a plain wide button. We will present a laboratory experiment in which approximately 30 participants are randomly assigned to an app-based smartphone survey with one of these three conditions: (1) radio buttons or check boxes (with tappable text), (2) radio buttons or check boxes embedded into wide buttons (which we call hybrid buttons), and (3) plain wide buttons. The questionnaire contains 23 choose-one questions as well as 3 choose-all-that-apply questions, most of which are adapted from the World Values Survey. We will evaluate performance based on completion time, self-reported ease of completion, and errors made when tapping a response option as well as the number of answer categories selected for the choose-all questions. We expect that hybrid buttons and plain wide buttons will improve performance compared to radio buttons and check boxes by providing a cue to respondents that they can tap anywhere in the row to activate it, thereby

increasing the perceived size of the response options. We also expect plain buttons to outperform hybrid buttons because they do not contain a smaller target that can compete for respondents' attention.

Optimal response formatting for fixed-field data items

Ivonne Figueroa

People are increasingly using their mobile devices to pay bills, order food, and to complete Web surveys, all of which require entry of numeric strings of data like telephone numbers and credit cards. Yet, mobile application designers are often unclear about how data entry should be formatted, particularly because various formatting forms exist. This presentation compares two format designs for fixed-field format data entries. One format design, auto-tabbing, is when separate fields are available for each "chunk" of a string entry. Another format design, masking, offers a format embedded within the field once the user activates the field. Little information is available about which design is more effective and optimal for fixed-field formatted entry. While usage of auto-tabbing formats may confuse users because they are not expecting to automatically advance to the next field, masking formats may slow performance because users intend to enter hyphens and special characters manually. In the present study, we are to empirically test whether users enter data more accurately, faster and with better satisfaction with auto-tabbed fields or fields with masking formatting. Participants will complete either four auto-tabbed or four masking formatted entry tasks in counterbalanced order. The tasks will require participants to enter information stored in memory including their phone number and date of birth, and information provided to them including a credit card number and login credentials. A satisfaction rating will be collected following each task. Other metrics collected will be time on task, errors, string responses (i.e., backspaces, special characters entered, etc.), and a final response entry. Implications of the findings will be discussed.

Dropdown response options in mobile surveys

Elizabeth Nichols

Dropdown response fields or "pick-lists" are a type of automated design where the user can click or touch on a single field to bring up a list of possible choices. For long lists, the choices are in a scrollable frame. When the user selects a choice, the list disappears, and that choice is displayed back in the single field on the screen. Dropdowns are often used in online surveys where there are space constraints or when there is a long list of response options, such as when someone selects a state in the U.S. On a PC, the design functions the same on different operating systems. However, on smartphones, the design functions somewhat differently depending on the phone's operating system. In the iOS system, the list displays at the bottom of the screen with a spinner that only shows a few possible choices at a time; while in the other operating systems, such as Android, the list functions more similarly to the way a PC would handle these fields. Qualitative evidence from mobile usability tests suggests that participants have difficulty using the iOS spinner. This talk will present findings from an empirical experiment comparing three possible designs for mobile surveys: the iOS spinner design, the Android pick-list design, and a radio-button design. Each participant will be randomly assigned to one of the three treatments, with 20 participants in each treatment. Participants' performance will be assessed with time-on-task, measurement error, and satisfaction. We expect that the participants in the iOS design group will take longer to complete the survey, be less satisfied with the design, and enter more incorrect responses than the other two designs.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Issues Related to Surveying and Interviewing Hispanics in the United States (Panel)

Issues Related to Surveying and Interviewing Hispanics in the United States

Anna Sandoval Giron *U.S. Census Bureau*
Lucia C. Lykke; Gerson D. Morales
Lucia Lykke *U.S. Census Bureau*
Gerson Morales *U.S. Census Bureau*
Ilana Ventura *NORC at the Univ of Chicago - Chicago, IL*
Rene Bautista-Martinez *NORC at the Univ of Chicago*
Rosa Avila *AcademyHealth/NCHS Health Policy Fellow*
Anna Sandoval *US Census Bureau*
Ana Gonzalez-Barrera *Pew Research Center*
Mark Lopez *Pew Hispanic Ctr.*

The latest estimates from the U.S. Census Bureau put the Hispanic/Latino origin population at 56.6 million (as of July 1, 2015), making it the nation's largest ethnic or racial minority. Nonetheless, until recently the unique challenges of working with this growing population have remained outside the mainstream survey methodology conversations. For instance Hispanics have often been lumped as one homogenous group. Nonetheless, Hispanics are a heterogeneous group in terms of language, ethnic identity, migration experiences, and voting patterns. This panel brings together a group of researchers who explore a variety of factors that impact work related to surveying and researching the growing Hispanic population in the United States. Lucia Lykke and Gerson Morales will present an examination of English use and its implications for data quality in cognitive testing of a Spanish translation of a U.S. housing survey. Ilana Ventura and colleagues will present on panel recruitment for Spanish speaking populations using AmeriSpeak as a case study. Rosa Avila will present a discussion on the methodological challenges to consider when analyzing survey data to evaluate health-related disparities between Latinos and non-Hispanic white subpopulations. Anna Sandoval Girón will present on Central and South American Indigenous respondents navigating racial identity categories in U.S. census forms. Finally, Ana Gonzalez-Barrera and Mark Hugo Lopez discuss the role that immigration and intermarriage rates play in reshaping U.S. Hispanic identity. Much like the Hispanic community in the United States, the panelists come from a variety of perspectives and seek to provide the breadth and depth of the Hispanic experience in the United States.

How much is too much? English Use and its Implications for Data Quality in Cognitive Testing of a Spanish Translation of a U.S. Housing Survey.

Lucia C. Lykke; Gerson D. Morales

Examining whether testing survey materials with bilingual vs. monolingual respondents is effective is a recent area of interest in multilingual survey methodology. Studies on this topic investigate whether and how bilinguals and monolinguals differ in their understanding of the cognitive interview task (Goerman 2006), their preference for taking a survey in one language or another (e.g., Trussell 2013), or compare understanding of key survey concepts between monolinguals and bilinguals (Goerman et al. 2016; Park et al. 2014). Although the research shows that bilinguals can help to uncover survey translation issues as

well as monolinguals in some cases, research is sparse regarding language proficiency of bilinguals in both English and the target language and how it might affect data richness and quality in cognitive interviews. Self-reported language proficiency can be an invalid and unreliable measure, and participants are rarely asked about their language preferences and non-English language abilities when recruited for studies. In this study, we first review language screener questions used in multiple studies conducted by the Census Bureau as criteria for selecting bilingual respondents. Second, we use data collected from 29 cognitive interviews with bilingual and monolingual Spanish speakers to test a newly translated housing survey to examine whether and how participants use English in a non-English interview setting, and how this compares to participants' language abilities in both languages. We examine the frequency of English use by bilingual Spanish speakers and the context of this English use. We compare English use by respondents during the interview with their self-reported English language abilities both before and after the interview, and with their post-interview stated preferences for survey language. We consider the implications of using bilinguals of varying levels of English and Spanish proficiency for testing survey questions, with future implications for the recruitment of bilingual cognitive interview participants.

Panel Recruitment for Spanish Speaking Populations: The AmeriSpeak Case Study

Ilana Ventura; Rene Bautista; David Gleicher; Carolina Milesi; Erlina Hendarwan

Obtaining representative samples of the Hispanic and Latino population is challenging in the survey research environment in the United States. This paper discusses some of the initial challenges and proposed solutions in obtaining a representative sample of Spanish-speaking households for a probability-based panel. To improve the representativeness of the Spanish-speaking segment of NORC's AmeriSpeak® Panel, NORC performed a qualitative assessment of the AmeriSpeak Spanish panel recruitment materials, including in-depth interviews and focus groups, to identify adjustments to the panel recruitment protocol. This paper discusses our methodology for the qualitative study performed, as well as our recommendations for improving the effectiveness of NORC's panel recruitment with Spanish speakers. The objective of this research is to provide insights to develop improved recruitment materials for Spanish speaking populations. AmeriSpeak is a nationally representative sample of U.S. adults, where households are selected randomly from NORC's National Sample Frame, which provides sample coverage for over 97 percent of U.S. households and includes additional coverage of hard-to-survey population segments, such as rural and low-income households. Recruitment is a two-stage process. For the initial recruitment, sample units are invited to join AmeriSpeak online by visiting the panel website or by telephone (in-bound/outbound supported). In the second stage, non-response follow-up is performed by way of field interviewers using face-to-face contacts with non-responders. English and Spanish languages are supported for online, telephone, and in-person recruitment.

Methodological Challenges When Analyzing Latino Health Disparities

Rosa Avila

Latinos are the fastest growing minority group in the US, and the need for monitoring and evaluating health disparities among Latinos is an important priority in public health. The purpose of this presentation is to highlight challenges and important issues to consider when analyzing survey data to evaluate health-related disparities between Latinos and non-Hispanic white subpopulations. Examples from two studies conducted with National Health Interview Survey (NHIS) data will be used to outline the methodological process of evaluating changes in disparities over time. The NHIS is a household interview survey on the health of the noninstitutionalized US population. The first study evaluated racial/ethnic health disparities at the national level since the Affordable Care Act, and the second study

used state-level data to evaluate the impact of the Massachusetts health reform on racial/ethnic health disparities. As the methods and results are discussed for these two studies, I will highlight limitations of the data and issues to keep in mind when designing such studies, analyzing disparities data, and interpreting results. Some of these issues include the comparability of data from translated questionnaires, the value of comparing subgroups of Latinos, pooling of data, over-controlling models, and power limitations. When conducting trends analysis, displaying disparity results that are intuitive to a nonstatistical audience can be difficult but possible. From this presentation, I hope to convey important information to consider when evaluating racial/ethnic disparities over time in the US.

Central and South American Indigenous, American Indian or Hispanic/Latino Respondents? Navigating racial identity categories in U.S. Census Forms.

Anna Sandoval Girón

According to the Office of Management and Budget (OMB), “American Indian or Alaska Native” refers to a person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment. Thus, individuals who identify as belonging to any of the Central and South American Indigenous groups should fit into the “American Indian or Alaska Native” category. However, these individuals may fit into this category without maintaining political ties such as tribal affiliation. Initial research by the language and cross-cultural research group at the Census Bureau on American Indian and Alaska Native questions regarding racial identification and tribal enrollment found that rather than identifying with “tribes,” Central and South American indigenous people tended to identify as Hispanics. Although they identified with their indigenous culture and language, they saw tribes as exclusively pertaining to indigenous groups of the United States (Schwede, Terry, & Fernandez, 2016). This poses a challenge that may lead to underreporting indigenous populations in the US, as these respondents may not see their identities represented in the American Indian race category, and the concept of tribal enrollment may be inapplicable to Central and South American indigenous. This paper has several goals: to discuss patterns observed in cognitive interviews; to review the census forms of other countries with large indigenous populations living in the U.S. (Mexico, Guatemala, Peru, and Bolivia); analyze indigenous identifications in these countries; and to discuss the impact that variations in indigenous identifications may have for how respondents interpret the U.S. Census American Indian or Alaska Native items. To conclude we will include recommendations on ways to better survey and capture indigenous central and South Americans.

Declining Immigration and high intermarriage rates are reshaping U.S. Hispanic identity

Ana Gonzalez-Barrera; Mark Hugo Lopez

The nation’s Hispanic population numbers 57 million and is projected to make up about one-quarter of all Americans by 2065. When reporting statistics about the U.S. Latino community, surveys from the U.S. Census Bureau and elsewhere rely on Americans to self-identify whether they identify with the group. Americans are also asked to self-identify their race. As a result, there are some Americans who have a Hispanic background or ancestry, but say they are not Hispanic. A recent Pew Research Center survey suggests that 5 million U.S. adults are self-identified non-Hispanics with Hispanic ancestry. This number is likely to grow as immigration from Latin America slows and intermarriage rates remain high. All this has implications for the future of Hispanic identity in the U.S. – how it will evolve and what terms might future Hispanic generations use to describe themselves. It also has implications for Hispanic population growth in the U.S. Population projections in recent years have emphasized the size and speed of Hispanic population growth. Yet if many who are of this ancestry do not consider themselves Hispanic,

that may impact the future of Hispanic population growth. This paper will compare basic characteristics, views of identity and attitudes about social issues of self-identified Hispanics and those with an ancestry but who do not self-identify as Hispanic. The analysis will be based on two surveys conducted by the Pew Research Center. The first is a ground-breaking survey of U.S. adults who say they have a Hispanic ancestry, but do not consider themselves Hispanic or Latino. The survey of 401 U.S. adults was conducted between November 11, 2015 and February 7, 2016. The second is a national survey of 1,500 self-identified Hispanic adults conducted from October 21 through November 30, 2015.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

2016 Pre-election Polling: Methods and Accuracy in Context (Panel)

2016 Pre-Election Polling: Methods and Accuracy in Context

Chase Harrison *Harvard University*
Natalie Jackson
Claire Durand *Univ. De Montreal*
Natalie Jackson *Huffington Post / POLLSTER.COM*
Ariel Edwards-Levy *Huffington Post*
Richard Clark *Castleton State College*
Jennifer Dineen *University of Connecticut*
Chase Harrison *Harvard University*
Andrew Smith *University of New Hampshire*
Zachary Azem *University of New Hampshire*
Stephanie Marken *Gallup Organization*
Jeffrey Jones *Gallup Organization*
Lydia Saad *Gallup Organization*
Jennifer Marlar *Gallup Organization*
Zachary Auter *Gallup Organization*
Frank Newport *Gallup Organization*

As we sit in the wake of the 2016 Presidential election and search for an explanation as to how so many public opinion polls could have been so wrong in their characterization of the election, many consider the impact of evaluating, comparing and aggregating data from an increasingly wide variety of methodologies. Polling accuracy had been an important research topic since the late 1940's as the industry looked for ways to avoid another incident like the missed prediction of the 1948 election. The topic became prominent again in 1997 after as a result of accusations by some commentators that the 1996 US presidential election had been a disaster for pollsters exceeding the magnitude of 1948. Although many 2012 pre-election polls underestimated the Democratic voter share, the most prominent poll aggregators correctly predicted the race. In 2014 both pollsters and aggregators overestimated the Democratic vote share. As a result, the 2016 AAPOR Annual Conference contained numerous papers and panels investigating the effects of various methodologies – especially methods of data collection and sampling - on the effects of pre-election outcomes. In this panel, we will continue and extend the discussions we began after the 2014 election cycle by attempting to understand forecasting errors in 2016 polls in light of potential lessons learned across multiple contexts. This diverse set of papers will explore the relationship between pre-election polling methods and accuracy across a variety of years, countries, races, electoral contexts, and survey methods.

More Polls, More Problems? 2016 Poll Proliferation and Accuracy

Natalie Jackson

There are more pre-election polls than ever, but that might not be a good thing. In the aftermath of the 2016 general election polling miss (mostly at the state level) there are more questions than ever about polling accuracy, quality, and methodology. We use the 2016 pre-election polls from our HuffPost Pollster database to investigate how polling in the aggregate has been affected by poll proliferation, and

in particular, how accuracy of our estimates was affected by different types of polls. We will highlight new developments, such as online 50-state polls, and how those affected the overall view of polling in 2016. Our analysis will compare polls from the last two U.S. presidential elections and attempt to explain why there has been a rise in polls conducted, who these new pollsters are, if the poll modes have changed and how these polls are affecting the aggregate estimate of our model. We'll also look at whether choices about excluding polls from the aggregate affected results. We will conclude by discussing the implications of these changes and what it means for this cycle and the future of polling.

Simply Unpredictable: The Relationship between Methodology and Bias in Pre-Election Vote Share Estimates

Jennifer Dineen

The horse race is the lifeblood of election news in the United States as news coverage typically focuses on which candidates are in the lead and which are falling behind. After a series of elections where bias in pre-election polls, especially at the national level, was largely unpredictable, many turned to poll aggregator websites for predictions. While poll aggregators were successful in 2008 and 2012, the 2014 mid-term and the 2016 general elections clearly demonstrated that aggregating pre-election polls does not always result in better, more accurate predictions than those produced by the individual polls that go into their formulae. Most importantly, in many states, poll averages are driven almost exclusively by either low cost IVR polls and/or polls conducted by partisan political polling groups. This brings us back to the need for better understanding of the implications that choices in method of data collection and sampling frame have on the estimates produced by pre-election polls. This paper examines the content of poll aggregation sites spanning the 2006 through 2016 election cycles. We examine state-level pre-election polls done during the final month of presidential, gubernatorial and competitive senate campaigns, categorizing each poll by method of data collection, sampling frame, organization type, and accuracy to understand the effect of these variables on the vote share estimates. Preliminary analysis indicates a relationship between method of data collection and the accuracy of the vote share estimate.

Comparing 2016 Election Results from Traditional Phone Studies with Web-based Methodologies

Stephanie Marken

Declining response rates and increased costs associated with traditional sampling and data collection approaches have led many researchers to more rigorously explore alternative methodologies, and web-based methodologies in particular. Many researchers turned again to these methods in the 2016 presidential race. In the final week preceding the 2016 Presidential vote, Gallup conducted an experiment comparing these new, web-based methodologies, with its traditional phone survey methodology. Gallup's experiment was designed to compare data from several data collection approaches, including a phone survey utilizing a random-digit-dialing (RDD) dual frame sampling approach and web-based methodologies. The comparisons allowed Gallup to explore attitudinal differences about the election and likely voter models between different methodologies and non-representative and representative sampling frames, allowing researchers to identify bias associated with different sampling frames. In this presentation, Gallup will provide details about the accuracy of these approaches when compared with the final election results. Gallup will also share data about the actual turnout of opt-in panel members, comparing their self-report and likely voter model prediction with their actual turnout based on state election board records.

Pre-Election Polling and Sampling Frame Decisions: a Case Study in Vermont

Rich Clark

Pre-election polling can provide tremendous insight into the public's motivation to vote, helping political scientists understand the voting public beyond what can be gleaned from voting results. For instance, pre-election polls provide insight into the issues that are important to voters, the relationships between issue positions and candidate support, and the extent to which split ticket voting is likely and by whom. But pre-election polling at the state level needs to reflect a sub-population of the state—those who will actually vote in the election. What makes pre-election polling more difficult than general population polling is that the target population—voters—does not yet exist at the time of the poll. In order to generalize poll results to this yet-to-be-known population of voters, pollsters choose sampling frames that will give them the greatest odds of finding real voters and excluding those who will not vote. The two most often employed frames are (a) voter registration lists and (b) dual-frame cell phone and landline samples. While voter lists only include registered voters and come with meta-data on voting history, they run the risk of excluding new voters and sometimes exclude phone contact information or included out-of-date phone contact information. On the other hand, the cell phone and landline dual-frames run the risk of including nonregistered voters and exclude information about voting history, relying instead on self-reports. This paper compares two very similar studies in Vermont, conducted closely in time, each employing different sampling frames. Using election results and demographic data, the analysis will explore the relative representativeness of the voting population from unweighted data from samples as well as explore the options available to pollsters in determining likely voters from each sample frame.

The Impact of Polling Methods on Estimation of the Vote in a Comparative Perspective

Claire Durand

Polling methodology has been under high pressure lately. First, there is a diversification of methods and new methods cannot always be fully tested. Second, pollsters have to rely on samples that are far from optimal. Third, they face different technological, socio-political and legal environments in different countries that restrain their possibilities and makes it difficult to use exactly the same methods in different countries. And finally, there were many elections and referendums where the two “sides” were virtually at par, which is the worst situation in terms of margins of error. This paper focuses on the impact of methods of administration on the estimation of the vote. It adopts a comparative perspective to better understand what happened in the US presidential election of 2016 and what it can teach us for the future. In order to do this, it uses different statistical procedures – local regression, box & whiskers plots, linear regression, ANOVA – to examine the impact of methods in elections and referendums held in Canada, UK and the US in order to see whether it informs us on the likely biases of different methods and on the circumstances that are more likely to lead to these biases.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Methodological Brief: Web Panel Recruitment, Retention and Response
Contrasting the Effect of Router- versus Email-Based Recruitment on Invitation Response to Online Surveys

Vanessa Boudewyns *RTI International*
Vanessa Boudewyns *RTI International*
Sylvia Tan *RTI International*
Kevin R. Betts *U.S. Food and Drug Administration*
Kathryn J. Aikin *U.S. Food and Drug Administration*
Claudia Squire *RTI International*

When conducting online surveys with convenience samples recruited through panel vendors, researchers have the option to invite panelists through either traditional e-mail or a router-based method. The router is an automated tool that uses real-time traffic to contact persons already engaged with the vendor's website/platform; hence, router invites may achieve higher response rates than e-mail approaches. However, very little is known about possible differences in response rates resulting from the two modes. We contrasted response rates of participants recruited via e-mail vs. router for a large-scale U.S.-based federally-funded web survey on consumer reactions to prescription drug advertisements. The target sample was adults with depression, insomnia, or high cholesterol, with representation from persons with minimal education, and from minorities. A total of 17,103 participants were invited, with 15,453 (90%) contacted via email and a supplemental 1,650 (10%) contacted via router. Router invites targeted more Non-white (42% versus 22% email), younger (10% versus 7%), Hispanic (28% versus 11%), male (49% versus 33%), and low education participants (43% versus 18%). ($p < .0001$, all comparisons). Overall, 53% of router invitees responded, versus 28% of email invitees ($p < .0001$). Logistic regression was used to model differences in response rates, while adjusting for demographic composition and possible interactions. Results revealed significantly higher baseline odds of responding to a router versus email invitation (AOR 1.35, 95%CI [1.12 – 1.63]). Patterns of response differed significantly between the two modes. More non-Whites (61% versus 47% Whites) and more Hispanics (68% versus 47% non-Hispanic) responded to the router invitation. This pattern was reversed among email invitees, where more Whites (30% versus 20% non-White) and more non-Hispanics (28% versus 21% Hispanic) responded (each adjusted interaction $p < .0001$). Supplementing with router-based recruitment could strategically target panelists ready to take a survey, and could boost response rates overall and from subgroups of interest, particularly Hispanics.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Methodological Brief: Web Panel Recruitment, Retention and Response

Increasing the response rate to a national online survey: what is the impact of adding reminder calls to an existing response maximisation strategy?

Sonia Whiteley *The Social Research Centre*

Email reminders, hard-copy letters and incentivisation are typical components of a response maximisation strategy for self-completion surveys. While reminder notices in general tend to increase response rates the effect on representativeness of the completed surveys is not always examined. Data from a national online survey of higher education students is used to assess the impact of the introduction of telephone reminder calls on both response rates and representativeness. Reminder calls were introduced to the suite of response maximisation activities to replace hard-copy letters which were becoming increasingly expensive and difficult to administer. Initial outcomes appeared promising with a dramatic increase in overall the response rate of eight percentage points. Despite the targeting of reminder calls to groups of students who are traditionally reluctant to take part, particularly young men, representativeness only improved marginally. Refinements to the current approach to response maximisation that will be trialled in future waves of data collection to improve representativeness are discussed.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Methodological Brief: Web Panel Recruitment, Retention and Response

Developing an Online Panel: Findings from Five Phases of UX Testing

Lauren A Walton *Nielsen*

Kay Ricci *Nielsen*

Tim Oltman *Nielsen - Oldsmar, FL*

Jennifer Hunsecker *Nielsen - Oldsmar, FL*

Carlos Jarava *Nielsen*

Jean Guerrettaz *The Nielsen Company*

Christina Eiginger *Nielsen*

Ally Glerum *Nielsen*

Usability testing best practices suggest using an iterative approach when evaluating respondent materials (Geisen and Romano Bergstrom 2016). Exploratory testing allows researchers flexibility to incorporate feedback throughout the design process, resulting in a finished product that most closely suits the needs of their respondents. This research presents the iterative process of designing and testing a new online panel recruitment webpage for Nielsen's online panel. Researchers conducted five in-lab usability tests from July 2015 through December 2016 on the American and German recruitment sites moving from low fidelity wireframes to production quality materials. The July 2015 Test 1 (n=16) examined the status quo of recruitment websites in English from multiple countries on a computer while collecting eye tracking data. Eye tracking, a technology that allows researchers to passively study people's eye movements, has been used mostly for questionnaire testing within the survey research field. Test 2 in December 2015 (n=23), tested static homepage images on iPads to collect feedback on initial new recruitment home page designs finding comments such as "What is this website representing?" Test 3 in February 2016 (n=20), put two design options head to head while collecting eye tracking data finding a preference for the second design exposed and a horizontal orientation for content. Test 4 May 2016 (n=75) was focused on Germany by collecting data on new prototype design versus existing materials. Not only was eye tracking data collected, neurological measures such as resonance and focus were evaluated. The final test in December 2016 will examine high fidelity materials using eye tracking and collect information on user experience on computer and mobile devices; testing will also take place in Spanish. This research will share findings from each study, the evolution of design in these recruitment materials, and demonstrate how to conduct iterative UX testing.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Methodological Brief: Web Panel Recruitment, Retention and Response
The Effectiveness of Providing a Self-Tracking Tool to Kids and Teens to Improve Daily Compliance in a Panel

Courtney Mooney *Nielsen*
Arianne Buckley *Nielsen*
Vicki Hoverman *Nielsen*

Nielsen developed a Portable People Meter (PPM) that can detect audio exposure to encoded radio signals. Participating household members ages 6 and over are asked to complete three tasks every day to measure their media exposure: (1) unplug their meter from the charger when they wake up, (2) wear their meter all day, and (3) charge their meter when they sleep. Nielsen uses monetary incentives to motivate panelists and improve cooperation rates. While these incentives have been effective at motivating adults, they have been less impactful among children and teens. In-depth interviews among former young panelists indicated that further engagement with the task of carrying the meter could help create good habits among this younger demographic group and increase cooperation rates. In Summer 2016, Nielsen developed a chart to enable young panelists to track their compliance. The purpose of the chart was to improve engagement with daily tasks and ultimately increase compliance rates. Additionally, parents are often burdened with the task of coaching their children to be compliant; the chart provided young panelists with a mechanism to track their own performance. Nielsen tested the compliance tracking chart over a four week period in Summer 2016 and in Fall 2016. Both tests evaluated the effectiveness of the chart on compliance rates among young panelists. The Fall 2016 test also offered a reward to panelists who met weekly compliance goals for all four weeks of the test. Analyses included compliance rate comparisons between pre- and chart-tracking periods. For the Fall test, the analysis also assessed whether a reward for meeting weekly goals further improved compliance. Results from both tests showed that having the ability to self-track their own cooperation with the daily task does improve compliance. Also, panelist feedback indicated that panelists enjoyed using the charts to see their progress each day.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Methodological Brief: Web Panel Recruitment, Retention and Response

How do changes in communications impact daily, research panel participation?

Adam Gluck *Nielsen*

Arianne Buckley *Nielsen*

Within Nielsen's Portable People Meter panel, each day panelists are asked to wear a small piece of hardware known as a meter. This meter detects inaudible codes in radio and television broadcasts, and this information is used to estimate audience sizes. Panelists are asked to wear their meter every day, however some do not. To encourage panelists to wear more often, Nielsen periodically offers extra monetary incentives to panelists if they wear during a specific period of time, usually a one to two week time period. Recently, Nielsen started testing changes to the content and timing of communications that are used to alert panelists to these special opportunities. The specific areas that Nielsen is testing are: Using incentive communications that have more imagery and less text. Currently Nielsen's communications contain more text than imagery, and the images are more decorative than informative. The research question that Nielsen looks to answer is whether or not using more imagery in rules explanations will increase panel participation during the special incentive periods and beyond. Notifying the panelists that they earned extra rewards shortly after the end of a special incentive period. Nielsen does not regularly notify panelists if they earned a "special bonus." The bonuses are simply added to their monthly incentive. Nielsen is developing communications to inform panelists that they earned a bonus during the special incentive period. The research question that Nielsen is seeking to answer is whether or not using these notifications result in higher daily panel participation both overall and during subsequent special incentive periods. Analyses of these initiatives will explore whether these efforts are a cost effective way to increase panel participation. The analysis will look at the impact during each particular incentive period, as well as the cumulative effect of these offerings over multiple incentive periods.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Methodological Brief: Web Panel Recruitment, Retention and Response

Who won't respond to your online probability panel?

Alexandra Brown *Federal Reserve Board of Governors*

Caitlin Eichten *Federal Reserve Board*

Understanding how online probability panels affect data quality is an area of growing research interest. To contribute to this research, our work uses the Federal Reserve Board's "Consumers' Use of Mobile Financial Services" (Mobile, <https://www.federalreserve.gov/econresdata/consumers-and-mobile-financial-services-report-201603.pdf>) survey to explore the connection between online probability panels, unit nonresponse and attrition. The Mobile survey was conducted for five years, with each subsequent year after the first containing both a new sample and a resampled group that can be analyzed as a panel dataset. Because this survey also contains information on those individuals who were invited to take the survey but did not start and those that started the survey but did not finish, we can explore both unit nonresponse and attrition. After a general exploration into who does not opt in when invited to join an online probability panel and a description of the respondents and non-respondents in the new and panel samples in our data, we employ a latent class analysis model to better understand the distinct groups of individuals that exist within our survey non-respondents. Our research will follow the work of Lugtig, Das, and Scherpenzeel (2014) . Understanding these distinct groups will contribute to the body of research surrounding nonresponse and attrition in online probability panel surveys. It will also help us understand the underlying biases that may exist and should be noted while preparing our reports. We plan to complete this work by early December.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Methodological Brief: Web Panel Recruitment, Retention and Response

Sink or Swim: The Case For (and Against) River Sampling as a Supplement to Online Panels Research

Steven A Snell *Qualtrics*

D. Sunshine Hillygus *Duke University*

David Vannette *Stanford University*

A growing literature evaluates the the quality of nonprobability-based online samples by benchmarking such data against data collected through probability-based online samples or through other survey modes; nevertheless, the preponderance of this research relies exclusively on data collected through the use of standing online panels, thereby ignoring river sampling techniques. This is an important omission especially because of known biases associated with regular or ``professional" panelists. For example, professional panelists tend to be more politically knowledgeable and engaged. We ask whether river sampling might mitigate the known biases of samples derived from online nonprobability-based panels. To this end, we compare demographically similar samples gathered through online nonprobability panels and through river sampling techniques on a variety of measures, including self-reported civic engagement and extremity of political attitudes, as well as data quality measures, including length of interview, length and quality of open-ended response, and frequency of straight-lining. Finally, we consider across a number of survey experiments whether panel or river respondents are more affected by informational stimuli. We arrive at a mixed set of findings: while river sample respondents are indeed less politically interested, they appear more hurried and less attentive. We conclude that the addition of river sample serves to reduce bias in estimates, but at the expense of precision, as river sample respondents produce noisier data.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Issues in Measurement: Classification, Proxies and Burden

Does Relation of Retrieval Pathways to Data Quality Differ by Self or Proxy Response Status?

Jinyoung Lee *University of Nebraska-Lincoln*

Robert F. Belli *University of Nebraska-Lincoln*

Household surveys often allow surrogate or proxy reports when sampled individuals are incapable of responding (e.g., health, language issues), not available or unwilling to take part in surveys, or minors. Collecting data from proxy respondents may help to reduce survey costs and survey errors due to unit nonresponse, whereas it also poses questions about data quality of proxy reports as well as methodological issues such as lack of control for self-selections. Albeit mixed, studies tend to find high levels of agreements between self and proxy reports or small differences in bias by response status (Bollinger & Hirsch, 2007; Moore, 1988). Meanwhile, other studies focusing on the cognitive response processes find that proxy respondents use different response strategies from those who report about themselves (Blair, Menon, & Bickart, 1991; Todorov & Kirchner, 2000). Using audit trails from the 2013 field test of the reengineered Survey of Income and Program Participation (SIPP), this study will address cognitive aspects of self and proxy responding. The reengineered SIPP has integrated an Event History Calendar (EHC) into the instrument, which allows respondents to navigate through the questionnaire across themes and times and index event histories in autobiographical memory. Respondents can use their own event histories as retrieval cues for reporting other thematically and temporally interrelated ones (Belli, 1998). This study focuses on cognitive retrieval processes, specifically on flexible memory searches during EHC interviews. The SIPP collects data from all members of the sampled households, while allowing proxy responding. Extending our previous research on EHC interviewing dynamics, navigation patterns in self and proxy interviews collected from the same respondents as well as between respondents will be compared by analyzing specific actions and their order of occurrences in completing EHC interviews to identify potential differences in retrieval pathways by response status and their relation to data quality.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Is It "Xenophobia" or Lack of Knowledge?

Change and diversity in Russian public opinion on USA

Natalie Rice *University of Tennessee*

Yuri Drakokhrust *Radio Free Europe/Radio Liberty*

Oleg Manaev *University of Tennessee*

Dean Rice *University of Tennessee*

Howard L Hall *University of Tennessee*

Michael Fitzgerald *University of Tennessee*

After Crimea annexation Russian anti-Americanism has significantly changed both in intensity and structure. Thus, according to Levada-Center, in July 2005 66% of Russians had positive attitudes towards USA, and only 26% negative ones. In January 2015 this correlation was 81% vs. 12% (n = 1600). How was it possible for the Russians to change their mind so quickly and so drastically? In order to answer this question, the authors performed statistical cluster analysis of the 2002 and 2012 Pew surveys, and used Chiozza' Dimensions of America theory (2010) as theoretical framework. USA, from this standpoint, is recognized by the foreign public as a multi-dimensional entity that can be loved and hated simultaneously, but rarely rejected all together. The data (2012) analysis produced three clusters of respondents according to their attitudes towards the U.S. The "consistently pro-American" cluster included 32% of the respondents who have positive attitudes to everything American: country, people, business, culture and technology. The "consistently anti-American" cluster, which included 42% of the respondents, was hostile to everything American. The third "intermediate" or "moderately anti-American cluster" included 26% of the people, who were quite fond of American business and culture, but still disliked American ideas being spread in Russia (n = 1000). It significantly differed from Belarus in June 2016: 42% vs. 23% vs. 35% (n = 1500). Comparative analysis of Pew surveys conducted in other countries proved that structure of Russians' attitudes to USA significantly differed from them. The data analysis provided a probable explanation of how it was possible for a majority of the Russian population to change their views drastically over a relatively short period of time. And its major mechanism was fundamental change in the attitude of the Russian public was due to a change of opinion among people of the "intermediate" cluster.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Is It "Xenophobia" or Lack of Knowledge?

Friends, Neighbors, Townspeople, and Parties: Explaining Canadian Attitudes toward Muslims

Timothy B. Gravelle *Wilfrid Laurier University*

The 2015 Canadian federal election campaign put into focus relations between Muslim communities in Canada and the wider Canadian society, featuring as it did debates around banning the niqab, a “barbaric cultural practices” hotline, and Canada’s commitment to accept Syrian refugees. At the same time, challenges in the relations between Muslims and majority-group Canadians were not a new development in 2015: they had in the past faced periodic strains due to criminal and terrorism-related events. The Canadian case is in fact reflective of a challenge in intergroup relations facing a number of Western democracies. In light of this, what accounts for majority-group Canadians’ attitudes toward Muslims in Canada? Drawing on data from the 2011 and 2015 Canadian Election Studies and theories linking outgroup perceptions to intergroup contact (friends), local demographic context at both the micro-level and meso-level (neighbors and townspeople), and political factors (parties), this paper seeks to explain why majority-group Canadians hold alternately positive or negative views Muslims in Canada.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

The Disgruntled Voter: Frustration, Anxiety and Candidate Popularity

In Search of a Political Anxiety Index: If You Come to a Fork in the Road, Take It

Harry L. Wilson *Roanoke College, Institute for Policy and Opinion Research*

David G. Taylor *Roanoke College, Institute for Policy and Opinion Research*

The 2016 election undoubtedly caused a great deal of anxiety for candidates, journalists, pollsters, and typical citizens. The election featured presidential candidates in both parties who tapped into what appeared to be a pervasive sense that the federal government was not working for the typical citizen. For more than a decade, we have also seen—and much has been written about—an increase in partisan polarization and a division within the populace. While these signals suggest an unease or anxiousness about government, we do not have an overarching indicator to measure that unease. We are all familiar with various measures of the feelings of consumers, and several of us even poll for that concept within our states. Those measures provide a periodic check that can be compared across jurisdictions as well as over time. While we have numerous measures of political sentiment, including trust in government, political efficacy, right direction/wrong track, favorable/unfavorable ratings, fear of crime, and others, we do not have an index that measures political anxiety. What follows is a first attempt by The Institute for Policy Opinion at Roanoke College to compile such a measure. We tested several combinations of questions during several polls conducted during and after the 2016 campaign. We checked for changes in the measure(s) and examined differences among several demographic groups. The challenge is to find a combination of questions that is relevant, somewhat timeless, varies over time in meaningful ways, and summarizes several concepts that factor into what a consensus of pollsters and political observers would recognize as “political anxiety.” We welcome comments, critiques, and suggestions, and we encourage others to build on this concept. We think this is an idea that has value, but one that will need to be fleshed out over time.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Issues in Measurement: Classification, Proxies and Burden

Objective and Perceived Burden: Which Survey Features and Respondent Characteristics Contribute to Both?

Robin Kaplan *Bureau of Labor Statistics*

Scott Fricker *Bureau of Labor Statistics*

Respondent burden is believed to have a negative impact on data quality, responses rates, and other important survey outcomes. However, burden is sometimes defined as the amount of time it takes to complete a survey (objective burden), versus respondents' subjective appraisals of how burdensome a survey is (perceived burden). Few surveys have measured both objective and perceived burden within-respondents in a single survey to determine what survey features and respondent characteristics contribute to each – the goal of the present study. Participants (N=1233) completed an online survey about how they spend their time. They filled out a time diary where they entered in all of their activities, along with their start and stop times, from the previous 24-hour period – a task requiring detailed thought and memory. They also answered stylized questions about their activities and how they spend their time in general. Afterward, they answered demographic questions, completed a social desirability scale, rated how burdensome, effortful, and sensitive they found the survey, as well as how well-rested they currently felt. We found that time spent on the time diary predicted perceived burden, but not the total time spent completing the survey. Ratings of how effortful participants found the survey, and how sensitive they found the survey topic, were associated with increased perceived burden. In contrast, level of interest, how well-rested participants felt, and social desirability were associated with decreased perceived burden. Females also reported less burden than males. Objective burden (time to complete the survey) was positively associated with time spent on the time diary, how effortful participants found the survey, and age. We discuss the implications of these findings for furthering our understanding of the survey features (e.g., highly detailed tasks) and previously unexplored respondent characteristics (e.g., social desirability, fatigue) that may influence both objective and perceived burden.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Issues in Measurement: Classification, Proxies and Burden

Are you really who you say you are? Two case studies exploring respondent-reported misclassification.

Kenneth M Pick *USDA National Agricultural Statistics Service*

Sarah Goodale *USDA National Agricultural Statistics Service*

Audra Zakzeski *USDA National Agricultural Statistics Service*

For surveys used to estimate population totals, whether sample units are correctly classified as in-scope or out-of-scope for the survey can have a sizable impact. This paper explores approaches National Agricultural Statistics Service (NASS) has taken to understand and measure respondent-reported misclassification. Respondent-reported misclassification can have a substantial impact on a survey program and the statistical estimates produced from it. For example, the wording of screening questions and how they are administered can determine whether a respondent qualifies or does not qualify for a survey. If a substantial number of respondents are misclassified, population estimates may be severely misstated. To study respondent-reported misclassification, NASS conducted follow-up misclassification surveys for two separate programs: one for the 2015 Certified Organic Survey and one for the 2015 Local Foods Marketing Practices Survey. For these two misclassification surveys, NASS survey methodologists deconstructed the existing surveys' multi-construct screening questions used to determine whether the sampled unit is in-scope for the survey, into simpler, single-construct screening questions. Telephone enumerators at NASS call centers then recontacted a subsample of the original surveys' samples and determined their classification with the new screening questions. Respondents who changed their classification were considered misclassified in the original survey. Although the primary goal for both misclassification surveys was to produce misclassification weights for the original surveys' data, the secondary goal was to understand respondent-reported misclassification in order to reduce it during future data collections. Data analysis of the misclassification data from both surveys, in corroboration with anecdotal evidence from interviewer-behavior coding, found that respondents appeared to have an easier time comprehending the enumerated questions that focused on simpler, single-construct screening questions. As such, this paper concludes with lessons learned from administering these two misclassification surveys and a discussion of how these lessons can be applied in order to reduce respondent-reported misclassification.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Shaping American Society: Effects of Gender, Age, Sexuality and Power

Factors Underlying Differences in Public Awareness of Elder Financial Exploitation

Melissa J. K. Howe *NORC at the University of Chicago*

Callan Jaress *NORC at the University of Chicago*

Kelly Pudelek *NORC at the University of Chicago*

Angela Fontes *NORC at the University of Chicago*

Bernard Dugoni *NORC at the University of Chicago*

Hannah Breslau *NORC at the University of Chicago*

Elder financial exploitation (EFE) is one of the most common types of elder mistreatment in the United States and threatens older adults' physical, psychological and financial health and well-being. Yet preliminary analyses of original, nationally representative, survey data (N = 1,042) suggest that many Americans (ages 18 and older) do not know how to detect, prevent or intervene in EFE. In this paper, we analyze AmeriSpeak Omnibus data collected in collaboration with NORC at the University of Chicago in December 2015 to provide baseline measures of public awareness of EFE and to illuminate how public awareness of EFE varies within the population. Measures reveal, for example, public assumptions regarding who is most likely to perpetrate EFE and what types of EFE people have observed first-hand in their families and communities. Overall, we find that 41% of respondents know someone who has been a victim of theft of money or property, 38% know someone who has been a victim of scams, 27% know a victim of identity theft, and 22% know a victim of power of attorney abuse. Yet little more than half of Americans (56%) indicate that they would know whom to contact if they suspected EFE; and only 31% would contact Adult Protective Services (APS) or another social service agency about suspected EFE. Our findings thus suggest clear avenues for improving public awareness of how to detect, prevent and intervene in EFE. To illuminate which Americans are in greatest need of such information, we test for statistically significant associations for age, education, income, financial literacy, race/ethnicity, gender, and other risk factors identified by experts of EFE. Finally, we discuss implications of our findings for intervention design, including suggested adaptations to promising approaches already under way.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Designing & Formatting Questions to Reduce Measurement Error

Loop-de-loos: Examining Respondent Reporting on Looping Questions

Antje Kirchner *RTI International*

Emilia Peytcheva *RTI International*

Shauna Yates *RTI International*

Ashley Wilson *RTI International*

Lesa Caves *RTI International*

Natasha Janson *RTI International*

Rebecca Powell *RTI International*

Respondents in surveys are often asked to respond to a series of follow-up questions that are repeated based on their response to filter questions (loops). For example, obtaining details about each employer a respondent has had. To determine the number of times a respondent goes through the loop, researchers can use one of two formats: (1) 'how many' or (2) 'go-again'. The 'how many' format asks respondents to report the number of occurrences followed by questions asking details of each occurrence. The 'go-again' format asks respondents to start with the first (or last) occurrence followed by more detailed questions. After answering the follow-up questions, respondents are asked if they have any other occurrences. If "yes", they continue to iterate through the loops. Such a task can become burdensome for respondents in either format, especially as the number of occurrences increases, potentially threatening data quality. This paper examines which loop format provides better data quality drawing on theories of motivated underreporting and research on reporting frequencies (e.g., Eckman and Kreuter 2015). We use data from the 2016/17 Baccalaureate and Beyond Longitudinal Study of college graduates (B&B, field test) where respondents were randomly assigned to one of the two loop formats. We evaluate the difference between loop formats in terms of number of reported occurrences, item nonresponse, breakoffs, and response time. Consistent with earlier research (Eckman et al. 2014; Eckman and Kreuter 2015) preliminary results suggest that data quality differs by loop format. Specifically, the reported number of employers is significantly higher in the how many format. Item nonresponse and breakoffs on the follow-up questions are significantly lower in the go-again format, and response times are significantly longer in go-again. We discuss the implications of our findings for data quality and the potential for imputation across the different formats.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

The Disgruntled Voter: Frustration, Anxiety and Candidate Popularity

The Frustrated Public: The Mood of the Electorate Ahead of the 2016 Presidential Election

Liz Kantor *NORC at the University of Chicago*

Dan Malato *NORC at the University of Chicago*

Marjorie Connelly *NORC at the University of Chicago*

Jennifer Benz *NORC at the University of Chicago*

Trevor Thompson *NORC at the University of Chicago*

Ahead of the 2016 presidential election, the public mood toward the candidates, the campaign, and the electoral system reflected a campaign season rife with negativity. A collection of studies conducted by The Associated Press-NORC Center for Public Affairs Research in 2016 reveal that, while majorities of Americans were interested and paying attention to the election, their feelings about it were largely downbeat, reflected in the election's low turnout. Key differences by age, education, partisanship, race, gender, and urbanicity emerge on many of these issues. The polls, conducted using the AmeriSpeak® panel, found over the course of the campaign that the public felt neither presidential candidate was capable of uniting the country, which Americans say is becoming increasingly divided on the most important values. Americans were frustrated, helpless, and angry about the election, seeing the disrespectful tone of political campaigns as surpassing the rudeness of everyday life. They expressed disappointment about the campaigns' focus on the candidates' personal characteristics rather than the issues that matter most to them, such as health care, Social Security, education, and terrorism. Further, the studies reveal that the negativity seen towards the candidates and the campaign also appears when looking at the political system as a whole, with many Americans saying that the two-party system is broken and unresponsive to the ideas of ordinary voters, and a significant minority expressing concerns about widespread voter fraud.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Issues in Measurement: Classification, Proxies and Burden

Latent Class Analysis of Worker Knowledge of Their Employment Status

Stanislav Kolenikov *Abt Associates*

Kelly Daley *Abt Associates*

Worker classification is a complicated and subtle area of the labor market. Inasmuch as workers have a choice between being an employee and being self-employed, it seems plausible that their choice will be affected by the perceived benefits and rights afforded by these two statuses. As increasing numbers of workers seek opportunities in the “gig” economy it is critical that they fully understand the rights and benefits they are giving up when moving into self-employment. The Worker Classification Survey, a national survey of 8503 adult workers, collected information on workers’ knowledge, specifically their self-perceived current job classification and the associated rights and benefits. The survey measures several concepts that are difficult for respondents to report accurately, including their own employment status as well as details about their employer, benefits, and tax forms. The Worker Survey also asks questions that some respondents may purposefully answer inaccurately, e.g., for those “working under the table” or not filing taxes on their income. Some of the responses in the survey, thus, are likely to contain measurement error. Latent class analysis (LCA) is a statistical technique to identify (latent) similar individuals in survey data. Past work has applied LCA to deal with measurement error in measurement of categorical variables such as illicit drug use, employment status, and mode effects. In this project, we utilize self-report of employment status, employer-treated status, and piloted series of labor classification questions which include measures of worker behavior control, job permanence, financial and employer relationships to identify the most common groups of workers. We show that the best-fitting model classifies 59% of the sample into employees, 24% into non-employee contractors, and 17% into workers that are difficult to classify unequivocally. The analysis also uncovers limitations of the instrument in determining the correct status as an employee or a contractor.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Designing & Formatting Questions to Reduce Measurement Error

The Impact of Working Memory on Response Order Effects

Beth Cochran *University of Nebraska - Lincoln*

Previous research in survey methodology has shown that primacy and recency response order effects do occur (Schuman & Presser, 1981; Sudman et al., 1996), and it has been theorized these response order effects occur when respondents do not extend enough cognitive processing to all of the response options (Krosnick & Alwin, 1987). In 1999, Knäuper conducted a meta-analysis determining that response order effects were more prominent among older adults than younger adults theorizing that some of the age effects could be explained by differences in working memory capacity. This current research extends the previous work on response order effects by examining the impact working memory has on respondents' response selection in attitudinal questions. Data for this study was collected during the fall of 2016, and using an address-based sample of adult residents in a Midwestern state, potential respondents were randomly assigned to an administration mode, either web or telephone. The data include working memory measures, the sentence span and alphabet span tasks, to assess the working memory capacity of respondents. Respondents are also asked to provide their preference on a variety of entertainment topics. Through randomizing the response options, the location of the response options (i.e., early vs. late) is analyzed separate from their content. The data is analyzed using logistic regression to determine if working memory capacity is a significant predictor of the response option location selected by a respondent, controlling for age and content of the response options. It is anticipated that working memory capacity will contribute significantly to primacy (web) and recency (telephone) response order effects.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Shaping American Society: Effects of Gender, Age, Sexuality and Power
Shining a Light on the Landscape of Online Harassment and Cyberstalking: The Findings of a National Survey of American Teenagers and Adults

Amanda Lenhart *AP-NORC Center for Public Affairs Research*

Michele Ybarra *Center for Innovative Public Health Research*

Kathryn Zickuhr *Data & Society Research Institute*

Myeshia Price-Feeney *Center for Innovative Public Health Research*

After a presidential campaign notable for its divisiveness and vitriol, both in person and on digital platforms, a deeper understanding of the landscape of online harassment, abuse and cyberstalking can help inform the national conversation around online discourse and harassment sparked by the events of campaign 2016. In recent years, media coverage has brought to light the ways in which new digital technologies appear to be enabling harassers to reach their victims in new ways. In the summer of 2016, the Data & Society Research Institute and the Center for Innovative Public Health Research undertook a 3,000-person nationally representative cellphone and landline survey of Americans age 15 and older in an effort to document the prevalence of digitally-mediated harassment, abuse and cyberstalking. The study examines 20 different modes of online harassment and abuse – including direct harassment tactics such as offensive name-calling, sexual harassment and physical threats; invasion of privacy, which includes exposure of sensitive personal information, hacking, monitoring and tracking, as well as the denial of access to digital accounts and platforms. The survey shows overall rates of abuse as well as who is most likely to experience each type of harassment and the relationship of the harasser to the victim. The study further examines the online climate and how witnessing and experiencing harassment relates to an individual's experience of being online, and looks at which witnesses are also bystanders who support and stand up for victims. The study asks victims of digital abuse not only about the harms they experienced but also the steps they take to protect themselves and mitigate some of the harassment directed at them.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Issues in Measurement: Classification, Proxies and Burden

How much do estimates of health insurance depend on who is asked to report from the household?

Aaron Maitland *National Center for Health Statistics*

Robin Cohen *National Center for Health Statistics*

Sarah Joestl *National Center for Health Statistics*

Health insurance coverage is an indicator of access to the U.S. health care system. The National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics (NCHS) has been an important source for measuring health insurance coverage in the U.S. population. The survey questionnaire and procedures are undergoing a redesign to improve the measurement of health topics, harmonize content with other federal health surveys, and reduce respondent burden. NCHS has proposed to revise the health insurance section in the new questionnaire to collect health insurance data on one sample adult and one sample child per household rather than asking these questions of entire families as has been done in the past. A simulation based on analysis of NHIS data from 2014 and 2015 found that estimates based on the sample adult data were only slightly different from estimates based on the family data. These estimates, however, were based on responses from the same family respondent, who may or may not have been the sample adult. This paper expands on this initial analysis by augmenting it with data from an experiment involving 2015 NHIS respondents that was designed to directly measure the impact of self versus proxy reporting on health insurance estimates. In it, the randomly-selected sample adult respondents were recontacted between one and three months after their initial interview and asked to complete a health insurance battery for the entire family. Those answers could then be compared to those given in the original NHIS interview. We directly compare how this change in survey design affects conclusions about health insurance. We also examine the consistency of reports about health insurance coverage between the two interviews when the respondent is the same and when the respondent is different.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Shaping American Society: Effects of Gender, Age, Sexuality and Power
The Causes and Consequences of Eroding Confidence in U.S. Institutions
Andrew Brett Dugan *Gallup*

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Designing & Formatting Questions to Reduce Measurement Error

Evaluation of the Impact of the Response Slider Scales: Validity, Cognitive Effort, and Moderation of Experimental Treatment Effects.

Sebastian Lundmark *Stanford University*

Flávio Azevedo *Universität zu Köln*

Jon Krosnick *Stanford University*

George E. Marcus *Williams College*

As the usage of web-based questionnaires have grown, so has the implementation of the so called 'slider scales'. In contrast to the more traditional 'radio buttons', the slider allows respondents to place themselves along a continuous scale without forcing them to mutually exclusive response categories. Instead, the slider is thought to allow respondents to more precisely place themselves along the theoretical attitudinal continuum intended to be measured. However, the potential benefits of sliders are still under scrutiny in the survey literature. While some find that they benefit measurement, others are more skeptical. To add to this literature, novel findings from previously published work as well as completely new experiments on the impact of slider scales compared to traditional ways of measuring attitudes will be presented. These experiments seek to compare the cognitive effort invested by respondents, if respondents translate the continuous sliders to the theoretical attitudinal continuum set out to be measured, and whether respondents meaningfully engage in the extra precision supposed to be given by the slider. In addition, these experiments investigate how the slider bar anchoring, the vertical or horizontal presentation of the slider, and the precision of the slider, affects experimental treatment effects, construct validity, and whether slider scales are more strongly moderated by respondents' cognitive ability and educational attainment compared to the more standard mutually exclusive radio buttons.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Designing & Formatting Questions to Reduce Measurement Error

Ordering Your Attention: Response Order Effects in Web-based Surveys

Frances M. Barlas *GfK*
Randall K. Thomas *GfK*
Nicole Buttermore *GfK*

Altering features of responses on a questionnaire, such as the order in which response options are presented, can substantively affect survey results. Moreover, response order effects have been found in a number of different modes of survey administration. While recency effects (where the last or near last response is more often selected) have been commonly found in interviewer-administered telephone surveys, some researchers have not been successful replicating these results. On the other hand, in self-administered surveys where response options are visually presented, like paper-pencil and online surveys, primacy effects (where the first or second response is more often selected) have been more common. While primacy effects appear to be more frequent in web-based surveys, such effects have also not consistently been replicated. In order to better understand the conditions under which order effects are most likely observed in online surveys, we collected together over 10 online studies representing data collected from about 100,000 respondents where response order was randomly assigned. These studies included questions that had scales varying from 2 to 11 categories, as well as scales presented vertically and horizontally. Measured dimensions included intensity (e.g., likelihood, usefulness, importance) and evaluation (e.g., good-bad, satisfied-dissatisfied). We generally found that when present, primacy effects were most likely to be demonstrated when scales were presented vertically rather than horizontally, more likely for unipolar rather than bipolar scales, and more likely with fully-labeled rather than end-labeled scales. A number of measures we looked at were associated with criterion measures, so we then examined how response order effects could affect measurement validity. This study shows researchers when response order effects could affect study results and provides guidance for better response scale construction.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Is It "Xenophobia" or Lack of Knowledge?

Do names matter?: Priming effect of Asian names on perceived suitability for political leadership positions

Hyungjin Gill *University of Wisconsin - Madison*

Asians are not seen as leaders in American society, as many merely see them as hard-working assistants. A national survey found that more Americans are uncomfortable voting for an Asian American for president than for a presidential candidate who is an African-American or a woman (Committee of 100, 2001). The six major Asian American ethnic groups in the U.S. that include Chinese and Japanese have surnames that clearly reflect their ethnicity, and past research suggest that ethnic names may activate negative racial stereotypes (Booth et al, 2012), which might have led some Asian people go through “whitening” of their profiles by adopting Anglicized names to avoid potential racial bias (Kang et al., 2016). Asian Americans have become the highest-income, best-educated racial group in the U.S. (Pew Research Center, 2013), and it is hoped that there would be more Asian involvement in leadership roles in American politics to shape future direction of the country. Based on the notion of priming (Domke et al., 1998; Roskos-Ewoldsen et al, 2002), this study tests whether implicit racial cues (i.e. ethnic names) embedded within unbiased, impartial political candidate speech transcripts trigger racial stereotypes in the assessments of leadership suitability. More specifically, this study conducts an experiment to investigate if the name variable, which has three levels (Asian name, Anglicized Asian name, and Western/English name), affects variation in perceived leadership. Each participant was given three non-racial, politically-balanced speech transcripts delivered by three different candidates – each transcript had speech deliverer's name (i.e. racial stimulus) written at the top. It is expected that racial priming will allow attitudes toward Asian Americans more accessible in memory, and that such attitudes will be applied during leadership evaluations. The insights gained from this study may assist in explaining the current American attitudes on race and leadership.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

The Disgruntled Voter: Frustration, Anxiety and Candidate Popularity

Anxiety and the 2016 US Presidential Election

Steffen Weiss *YouGov*

Samantha Luks *YouGov Polimetrix*

Rebecca Phillips *YouGov*

Anxiety and enthusiasm have been found to be motivating factors in the vote choice as part of the Affective Intelligence framework. This paper uses a two-fold approach to re-evaluate that link. We established a pre-election baseline measure of general respondent anxiety using the GAD-7 measure the levels of anxiety in the population close to election time, and to study what effect anxiety has on voter registration, turnout, and the vote choice. In addition to reported registration dates, official voter registration data will establish if anxiety was a motivating factor for voter registration and turnout. The survey also included a survey experiment where two of three subgroups were exposed to a one minute video treatment of each candidate to establish if respondents felt specific anxiety about the candidates, and if this anxiety was activated along party identification lines or if there were cross pressures between gender and party identification. A post-election recontact wave established what impact the election outcome had on respondent's baseline anxiety measures.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Designing & Formatting Questions to Reduce Measurement Error

Certainty and Accuracy: Question Order Effects in Factual Knowledge Estimation

David Vannette *Stanford University*

Bradford S Jones *University of California Davis*

Survey respondents are often asked factual questions and in some cases they may be asked to evaluate their certainty in the accuracy of their responses. However, a large body of research has demonstrated that the order of questions may affect the responses provided. This paper seeks to address the question of whether the order of asking factual questions and certainty questions about the accuracy of the answers to the factual questions may influence responses. This study evaluates a survey experiment aimed at assessing the effects of respondent evaluations of their certainty in their ability to estimate a factual statistic. Respondents were randomly assigned to either 1) prospectively evaluate their certainty in their ability to accurately estimate the number of Latino/as that have died on the U.S. side of the border while crossing from Mexico or 2) retrospectively evaluate their certainty about the accuracy of this same estimate after having provided an answer. The results indicate that there is an effect of the question order that researchers should be aware of when asking about factual knowledge. This research has implications for understanding respondent cognition and potentially for improving data quality.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Shaping American Society: Effects of Gender, Age, Sexuality and Power

Women in leadership: Why It Matters

Damla Ergun *Global Strategy Group*

Damla Ergun *Global Strategy Group*

Michael Smith *Global Strategy Group*

Andrew Baumann *Global Strategy Group*

Gender inequality is one of today's biggest challenges in the workplace. Currently, there are only 21 women CEOs of the Fortune 500 companies and 60 percent of 22,000 publicly traded companies have no women board members. To understand Americans' views of and personal experiences about gender equality in the workplace, Global Strategy Group, with a grant from the Rockefeller Foundation, conducted a 20-minute, national online survey among 1,011 U.S. adults, age 18 and older, in Spring 2016. The results show that, majorities of men and women alike, believe that women and men are equally qualified to lead businesses and say that it is highly important to them that men and women have the same opportunities for advancement. Nonetheless, Americans acknowledge the forces holding women back, saying that traditional workplace cultures as well as women's being seen as prioritizing family over work contribute to their lack of representation in leadership positions. Further, they see other concrete barriers for women pursuing leadership positions, including lack of support from mentors in securing top positions and for career advancement more generally, as well as lack of access to personal connections that men have which help with career development. Results highlight the unique importance of female role models—among those who have had mentors that supported them in the workplace, majorities of men and women alike say their mentor was the same gender as them. Specifically, among women who had mentors in the workplace that have supported them in their careers, nearly two-thirds (63%) say that their mentor was another woman, while more than three-quarters of men (77%) say they had male mentors. Given the scarcity of women in leadership positions in the workplace, this suggests that many women lack the very mentorship that provides the guidance that is essential to success in their careers.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Is It "Xenophobia" or Lack of Knowledge?

Brexit and attitudes to foreigners: The UK is not unique

M.D.R. Evans *University of Nevada*

Jonathan Kelley *University of Nevada*

Aversion to immigrants was a defining issue in political debates about Brexit and has grown into a major issue in many European Union nations. But if Britons are not uniquely hostile, then Britain's exit from the EU may be only the first symptom of enduring conflicts that will plague EU nations in future years. A well-established indicator of prejudice – aversion to outgroups as a neighbours – shows that prejudice against immigrants, other races, Muslims, Hindus, Jews, and Gypsies are all relatively low in Britain. This is as expected from the general decline of prejudice with socioeconomic development, leaving Britain similar to other prosperous EU and Anglophone nations. Confirmatory factor analysis suggests a single latent ethno-religious prejudice generates all these specific prejudices. Replication using other measures of prejudice and another cross-national dataset confirms these findings. Detailed analysis of Britain, Germany, the Netherlands, and the USA shows that prejudice in the EU declined to a minimum in the mid-90s, perhaps lower than elsewhere, but is now rising steadily, becoming higher than elsewhere. Data are from the pooled World and European Values Surveys (over 450,000 individuals, 300 surveys, and 100 nations for this analysis). Analysis is by descriptive, multi-level, and structural equation methods.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

The Disgruntled Voter: Frustration, Anxiety and Candidate Popularity

“Tracking Americans’ Images of the Candidates as the 2016 Campaign Progressed”

Frank Newport *Gallup*

Jeffrey Jones *Gallup*

Stephanie Marken *Gallup*

Lydia Saad *Gallup*

Gallup tracked the American public’s perceptions of the Democratic and Republican presidential candidates on a nightly basis from July 2015 through Election Day. During the pre-primary stages Gallup measured opinions of 16 Republican candidates and 5 Democratic candidates. These lists were winnowed as the primary phase of the campaign progressed, and ultimately reduced to Hillary Clinton and Donald Trump in June. Overall, Gallup assessed the candidates’ images based on more than 200,000 interviews. This paper will focus on the analysis of changes in the name identification and favorable/unfavorable images of the presidential candidates over time, providing unique insights into the ways in which Americans and party supporters responded to the candidates’ campaigns and key events in the nomination and general election contests. The continuing research allows for the examination of which campaign events, including debates and major news stories, did or did not change views towards the candidates. The paper will examine differences in the images and ultimately the success of candidates who had pre-existing name recognition and those who started being virtual unknowns. The paper will also detail the trajectories of Donald Trump and Hillary Clinton’s images as they moved towards gaining their party’s nominations, and what happened, or did not happen, to their images as the general election campaign progressed.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

The Disgruntled Voter: Frustration, Anxiety and Candidate Popularity

Proven Failure: First-Time Presidential Voting and Political Trust

Eunji Kim *University of Pennsylvania*

Jin Woo Kim *University of Pennsylvania*

Democratic theorists relish political participation not just for its intrinsic benefits in promoting community-mindedness among citizens, but also as an instrumental tool that fosters a sense of legitimacy and political efficacy. Yet to an extent that these benefits are theorized in a small-scale self-government setting, it is unclear whether participation in a representative democracy where many citizens participate only through voting has similar effects. We argue that voting typically results in eventual disappointment and frustration—either because one’s preferred candidate loses or because the president routinely overpromise and under-deliver—, which may have long-term consequences on political trust. Capitalizing on an exogenous variation in voting eligibility from age restrictions, we identify the downstream effects of first-time presidential voting on distrust. Drawing on the ANES time-series data (1974 to 2008), we find that being eligible to vote in a presidential election undercuts political trust 2 or 4 years down the road by several percentage points, a tendency that is more pronounced under a failing economy. We discuss implications of our finding for political socialization.

May 20th, 2017
1:45 PM - 3:15 PM
Concurrent Session H

Is It "Xenophobia" or Lack of Knowledge?

Remembering the Alamo: Demographic change and Texas politics

Tom Holub *University of California, Berkeley*

Texas, the second-largest state by population, is already a majority-minority state, and based on demographic projections supplied by the Office of the Texas State Demographer (OSD), Hispanics will soon outnumber whites in the state. Because Hispanics in Texas (and nationwide) vote for Democrats more frequently than Republicans, this demographic shift has serious implications for the expectation of Texas being a reliable “red” (Republican) state for Presidential elections. Over time, similar demographic trends have helped turn California from a mixed state which produced Ronald Reagan (and voted for him twice), to perhaps the most reliable “blue” state in the country, and also changed New Mexico from a historical toss-up state to a reliable blue state. My research aims to produce a robust population-based simulation predicting how the projected demographic changes in the Lone Star State will affect its leanings in future Presidential elections. I have built a Monte Carlo simulation in R to generate results for elections based on party preferences, voter turnout, and the effects of Voter ID laws for these ethnic groups, based on population projections from three different migration scenarios provided by the Texas OSD. An interactive demonstration of the simulation can be accessed at <http://research.inl.org/> The U.S. may be on the verge of a political shift as substantial as the one that switched the South from blue to red after the passage of the Civil Rights Act in 1964. In today’s increasingly multi-cultural United States, the xenophobic rhetoric of many mainstream Republicans is unlikely to woo the voters of the future. And while the idea of a permanent Democratic majority is as unrealistic now as the idea of a permanent Republican majority was in 1969 (Kevin Phillips), or in 2004 (Karl Rove), it is clear that the Republicans cannot win the presidency without winning Texas.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Effective Strategies for Conducting Surveys With Low Income Populations (Panel)

Effective Strategies for Conducting Surveys with Low-Income Populations

Sara Skidmore *Mathematica Policy Research*
Kim Mook and Sarah Forrestal
Deborah Herget *RTI International*
Kim Mook *Mathematica Policy Research*
Emily Weaver *Mathematica Policy Research*
Holly Matulewicz *Mathematica Policy Research*
Kathleen Feeney *Mathematica Policy Research*
Sarah Forrestal *Mathematica Policy Research*
Daniel Friend *Mathematica Policy Research*
Forest Crigler *Mathematica Policy Research*
Karen Donelan *Mongan Institute Health Policy Center*
Tiffany Waits *Mathematica Policy Research*
Myley Dang *Mathematica Policy Research*

Low-income individuals comprise a large proportion of respondents in studies conducted to inform public policy and evaluate government-sponsored programs. Conducting surveys with low-income populations presents unique challenges including difficulties locating respondents, respondents' literacy and internet access, coverage error, and approaches to training telephone and field interviewers. This population contains numerous distinct and diverse subpopulations, each with its own set of challenges that require special considerations to ensure participation in the research studies so often aimed to improve the well-being of lower income populations. This panel seeks to better understand the diversity within the low-income population, bringing together survey researchers experienced in surveying subgroups of the low-income population, including noncustodial fathers, low-income families with children, and individuals with significant disabilities. The panelists will discuss findings from studies of government-sponsored programs and explore many of the barriers faced by these groups and the challenges in gaining their participation in studies, including: challenges locating highly difficult to find respondents, considerations regarding data collection mode and its implications for reducing nonresponse bias, and engaging low income populations in web surveys. The panel also includes a meta-analysis, which provides a systematic review of research on data collection methods with low-income populations. Papers presented in this panel seek to broaden practitioners' understanding of low income populations, provide insight into conducting survey research within low income subpopulations, and discuss strategies and tools to overcome barriers to conducting research with this diverse population.

Commercial Locating Database Efficacy for Telephone Surveys of Low-Income Populations

Kim Mook and Sarah Forrestal

Low-income populations are often surveyed for the study of federal assistance programs. Locating identified program participants can be challenging due to mobility, variable employment status, and phone numbers that cycle in and out of service. To address locating challenges in a telephone survey of low-income households with children (n=11,496), we used two common commercial databases that

aggregate proprietary and public data sources about individuals and households. Project grantees provided sample frames constructed from either program administrative records or parent consent forms. We submitted households' contact information (name, address, and phone number) to the databases to obtain additional phone numbers prior to data collection. We compared 1) database hit rate to the frame, and 2) phone number source and quality, defined as successful respondent contact. We hypothesized that databases would vary in demographic representativeness and phone numbers provided by multiple sources would be better quality. Compared to one frame, preliminary results show databases were less likely to return hits on Hispanic and white heads of households but differences were not found by age or gender. Our investigation of the relationship between phone number quality and phone number source(s) (that is, the frame or a database) showed that phone numbers provided by multiple sources were more likely to be of good quality; phone numbers on the frames, regardless of the frame source, were more likely to be good than numbers provided by a database. Phone quality varied by source within a database, indicating sources should be prioritized. In this presentation, we will recommend guidelines for prioritizing contact information provided by multiple sources in surveys of low income households to maximize efficiency and sample representativeness.

Characteristics of Low-Income Fathers Who Complete by Phone vs. Field

Emily Weaver and Daniel Friend

Many human and social services exist to serve low-income fathers and families. Since 2005, the Office of Family Assistance has been authorized to spend ~\$150 million on programs that focus on healthy relationships and responsible fatherhood, many of which are geared towards low-income fathers. The Office of Child Support Enforcement has also established initiatives to help custodial and non-custodial fathers with their child support obligations. Several research projects have been and are being conducted to evaluate these government-funded programs. Along with this necessary program efficacy and effectiveness research comes the unique challenges of data collection with low-income fathers. Little guidance and research exists on choosing data collection methodology with this population, understanding the diversity within this group, and whether more expensive modes of data collection reduce nonresponse bias. The current research seeks to address this dearth by examining the response rates of two projects and comparing mode of completion. Both projects, sponsored by offices under the Department of Health and Human Services, evaluate social/human service programs for low-income fathers (Project 1 n= 5,522; Project 2 n=10,173). All sample members (N=15,695) completed a telephone baseline survey as they were recruited into the programs. A one-year follow-up survey was conducted by telephone with in-person field locating for hard-to-reach sample members. We will analyze the baseline demographic data (e.g., age, income, number of children) to examine their follow-up survey response rates and their mode preference at follow-up with the goal of identifying whether differences exist by race/ethnicity, age, and other demographic variables. The results from this research will help future projects think about research designs that improve coverage of this population and reduce nonresponse error, but also seek to better understand the diversity that exists within this low-income population of men and whether data collection methods need to better tailor to any subgroups.

An Operational View of the Digital Divide: Challenges and Opportunities in Engaging Low-Income Populations with Disabilities in Web Surveys

Holly Matulewicz, Karen Donelan, and Forest Crigler

Web surveys continue to attract interest, as they cost-effectively leverage the benefits of self-administration and computer-assisted interviewing (such as convenience of timing or pace of completion using programmed skip logic, text fills, and quality checks). Further, web surveys can

potentially eliminate contact barriers associated with phone or mail. Until recently, disparities in access, known as the digital divide, deterred researchers from using web surveys with low-income populations. Despite increasing access, limited information exists to inform planning. This study addresses that gap. First, what proportion of low-income individuals leave sufficient digital footprints to facilitate a viable email address match? Second, what proportion will participate by web and how many will do so across multiple rounds? Results are based on a multimodal panel survey of health plan enrollees (dualy eligible Medicare and Medicaid beneficiaries). The sample included 720 community-dwelling adults, ages 18-64, with significant physical or behavioral health disabilities. The survey was administered in two rounds, one year apart, with response rates of 63 and 60 percent, respectively. Initial contact information did not include emails. Using a search vendor, we matched at least one email for at least one-quarter of the sample members (25 percent in round one and 40 percent in round two) and many matches yielded multiple emails. Despite promising match rates, email invitations bounced back from 32 and 64 percent, respectively. Only a fraction of web respondents had matched emails, showing positive impact from the advance letter. Participation by mode was consistent across rounds and extremely low by web (70 percent completed by phone, 25 percent by mail, and 5 percent by web). Only 3 respondents completed by web in both rounds. Clearly, more work lies ahead to increase participation by web, including securing current email addresses, developing engaging, low-burden invitations, and providing supports for navigating online forms.

A Systematic Review of Data Collection Methods for Low-Income Populations

Kathleen Feeney, Daniel Friend, Tiffany Waits, and Myley Dang

According to the Government Accountability Office (2015), more than 80 federal programs are designed for low-income populations. Research examining the efficacy and effectiveness of these programs is also common. For example, the Administration for Children and Families, Office of Planning, Research, and Evaluation has funded contractual research on healthy relationship and responsible fatherhood programs geared toward low-income families for over 20 years. Despite the plethora of research on low-income populations, little formal guidance exists on best practices and challenges regarding data collection modes, issues related to measurement error, and important subpopulations that exist under the umbrella of low income populations, many of which have their own set of challenges and require special understanding culturally. Although some systematic review have been conducted on these topics, they have mainly focused on health and medical research, not human or social services. We will address this gap by conducting a systematic review and meta-analysis of data collection related issues pertaining to low-income populations. Our review will build upon existing systematic reviews by including research on social/human services conducted within the last 15 years. We will search several databases (e.g., PubMed, EBSCOhost, OVID, etc.) and gray literature (e.g., AAPOR conference abstracts, federal government reporting websites) to identify relevant studies involving low-income populations. These studies will be reviewed with an eye towards best practices regarding coverage of particular low income subpopulations, data collection modes, response rates (RR) and reducing nonresponse bias, and measurement error concerns. Our goal will be to identify the probable RR for each mode of data collection and the challenges and best practices associated with each. The result will be a compendium of potential data collection methods and tools to guide researchers in selecting data collection methods for low-income populations.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Principles and Case Studies for Innovation in Government Statistical Agencies (Panel)

Principles and Case Studies for Innovation in Government Statistical Agencies

John L. Eltinge *U.S. Census Bureau*

Thomas A. Louis, PhD, Johns Hopkins Bloomberg School of Public Health tlouis@jhu.edu

Thomas A. Louis *Johns Hopkins Bloomberg School of Public Health*

John M. Abowd *U.S. Census Bureau*

Linda J. Young *National Agricultural Statistics Service, USDA*

Greg Peterson *Statistics Canada*

Brian A. Harris-Kojetin *Committee on National Statistics/National Academies*

In keeping with the conference theme, large-scale statistical organizations have found it increasingly necessary to embrace change and diversity in stakeholder information needs, and in the data sources, methodology and technology used to address those needs. This session explores the innovation processes used to address these institutional challenges. Special attention is directed toward the integration of general innovation principles and practical illustrative examples, drawing on experiences from three U.S. statistical agencies and from Statistics Canada. The first paper highlights the needs for innovation in statistical agencies, and presents concrete proposals for changes in institutional culture, technology and bureaucratic processes. The second paper reviews four case studies, and related lessons learned, of large-scale innovations at the U.S. Census Bureau: the Small Area Income and Poverty Estimates Program, the Longitudinal Employer-Household Dynamics Program, the Management Organizational Practices Survey and the Microdata Analysis System. The third paper uses examples from the U.S. National Agricultural Statistics Service to explore factors that are of special importance in accelerating the transition from methodological research to full-scale production. In contrast with the first three papers, which draw on experiences from the decentralized U.S. statistical system, the fourth paper discusses innovation in Statistics Canada, which is a centralized national statistical office. This paper emphasizes creation of conditions that promote innovation throughout the organization, with illustrations from three cases involving, respectively, product innovation in a satellite account; process innovation based on modeling of non-survey data; and a second form of process innovation that uses crowdsourcing. The fifth paper uses three applications from the U.S. Bureau of Labor Statistics to emphasize practical benefits of statistical innovation; the use of empirical information to guide change management; navigation of issues related to standardization and process maturity; and alignment of individual and group incentives with measurable features of a proposed innovation.

Fostering Statistical Innovation in Federal Statistical Agencies

Thomas A. Louis, PhD, Johns Hopkins Bloomberg School of Public Health tlouis@jhu.edu

Innovation in and by Federal Statistical Agencies (FSAs) isn't optional, because these are competitive times and FSAs no longer have a monopoly on collecting, curating and disseminating demographic and economic information. Government surveys may continue to be the most trusted source of information, but without considerable innovation the relevance, accuracy, coverage, timeliness and cost-effectiveness of FSA collected and disseminated information will diminish. Issuing an "innovate" order by the Chief Statistician of the U.S. or other leader will have little effect without substantial cultural,

technological and bureaucratic change. Against this background, I discuss: - Cultural changes: including disrupting the culture of risk avoidance and resistance to change by amplifying the culture of experimentation, rewarding success and not punishing lack of full success of a speculative project; promoting a culture of peer review; increasing collaboration among the FSAs and with academe; encouraging researchers to aim high and submit to the leading survey-focused and broader domain journals. - Technological changes: including availability of high performance computing with sufficient storage, CPU capacity and speed to support development, evaluation and implementation of cutting edge procedures; flexible and timely software acquisition and upgrades. - Bureaucratic changes: including flattening the decision hierarchy, taming meta-work (work on work), providing "venture capital" to free up time and energy and go beyond attending to day-to-day operations, streamlining hiring and improving work conditions so that employment in an FSA is competitive with other opportunities.

Lessons Learned about Innovation in Federal Statistics

John M. Abowd U.S. Census Bureau john.maron.abowd@census.gov

Many factors contribute to the difficulty of innovation in the federal statistical system. Successes and failures warrant examination as case studies. This paper considers four such projects, all conducted at the Census Bureau. The Small Area Income and Poverty Estimates produces model-based estimates for states, counties and school districts. The program arose from a statutory mandate to disseminate such data without an accompanying budgetary supplement to enhance data collection. Since this scenario closely parallels the current environment, it is instructive to understand how the SAIPE system was developed and implemented. The first 21st century statistical system at the Bureau was the Longitudinal Employer-Household Dynamics Program, which produces local labor market data combining demographic and economic sources. Partnership with academic institutions and an engaged user community brought subject-matter experts and technical experts onto the same team. The goal was to use existing administrative, census and survey data rather than collect new data. A third example is the Management Organizational Practices Survey—first conducted as a supplement to the 2010 Annual Survey of Manufactures. In this project, also a partnership with academics and a specific user community, combining components of an existing survey mounted in many other countries with a regular Census Bureau survey allowed cost-effective simultaneous measurement of management practices and manufacturing outcomes. The final example is the Microdata Analysis System, which attempted to develop an automatic statistical system with direct access to confidential microdata. The goal of the system was to implement a restricted set of statistical analyses that would allow case and variable selection for a limited class of models. The experimental system was never implemented because the automated statistical disclosure limitation could not be done successfully. Not all innovations completely succeed, but all generate knowledge that improves the federal statistical system.

Moving Research into Production in a Federal Statistics Agency: Opportunities and Challenges

Linda J. Young, USDA National Agricultural Statistics Service, Washington DC 20250

Linda.Young@nass.usda.gov

Balancing the competing needs of adopting improved methods and technology and producing timely statistics is a constant challenge for Federal Statistical Agencies (FSAs). As is the case for other FSAs, USDA's National Agricultural Statistics Service (NASS) has evolved over time. However, once a good set of methods is in place, the tendency is to make minimal changes until a clear need is identified. Yet, today NASS is adopting new statistical methods and technology at an increasing rate due to several

factors. One is that NASS has identified some issues with its current methods that required major changes. As an example, in 2009, NASS conducted a study that identified substantial misclassification of land sampled from the NASS area frames; some farms were classified as non-farms, and some non-farms were classified as farms. At that time, for the Census of Agriculture, the adjustment for under-coverage of the NASS list frame assumed no misclassification. Resolving this issue prior to the 2012 Census of Agriculture led to major changes in the methodology for that census. Second, NASS is increasingly asked to report on emerging sectors of agriculture, such as organics, horticulture, and local foods. NASS's traditional methods, which are designed for production agriculture, are not sufficient for these specialty farms, which tend to be smaller, more transient, more diverse, and more dispersed than traditional agriculture. Identifying statistical methods that provide precise estimates for these studies has led to the adoption of a number of methodological changes. Third, the research staff are team members on projects with large statistical challenges, thereby facilitating the movement of the methods from research to production. This paper uses these and other examples to illustrate NASS's movement toward more rapid adoption of research methods as well as the challenges that have been and continue to be encountered.

Innovation in a Centralized National Statistical Office

Greg Peterson, Dir Gen, Agriculture, Energy, Environment, Statistics Canada greg.peterson@canada.ca

Innovation within a National Statistical Office (NSO) is essential to ensure that statistical programs remain effective, relevant and efficient. Creating conditions that promote innovation throughout the organization is key. The OECD Oslo Manual describes innovation under two broad categories: technological and non-technological. Technological innovation involves the introduction of new products or new processes. Non-technological innovation encompasses new organizational structures and new marketing techniques. Although this manual was developed to measure market-based goods and services, it is used in this paper as a model to describe innovation within Statistics Canada. By using the example of three recent projects within Statistics Canada, the paper will show how different types of innovation are linked. It will also show the conditions that Statistics Canada, as a centralized NSO, has put in place to encourage and support these types of innovative activities. The first of these projects include the development of a satellite account on clean technology as a form of product innovation. The second, an example of process innovation, is a project that saw the replacement of a mid-year survey of farm operators with a modelled approach using earth observation and agro-climatic data. The final example involves experiments in the use of crowdsourcing as a way of identifying building characteristics, as another example of process innovation.

Enhancement of Innovation in Large-Scale Statistical Organizations

John L. Eltinge, U.S. Census Bureau eltinge.john@gmail.com

This paper examines a range of factors that can be important in efforts to enhance innovation in government statistical agencies and other large-scale statistical organizations. Following a brief review of the general literature on adoption and diffusion of innovations (e.g., Rogers, 2003 and references cited therein) and of specific studies of innovation in government statistical agencies (e.g., Dillman, 1996; and National Research Council, 2011), we focus on four general ideas. First, we emphasize that innovation in statistical organizations is of interest primarily for the resulting practical benefits. These benefits may include creation of a fundamentally new statistical product; or substantial improvement of an existing product or process. Second, practical evaluation of those benefits may be very challenging, and decisions to initiate or continue a specific innovation effort often must rely on limited empirical information regarding the balance of quality, risk, cost and stakeholder-value profiles for those

products. Third, the abovementioned profiles may be heavily influenced by the degree of maturity and standardization encountered for the specific technology, methodology or data source under consideration. Finally, we highlight the importance of aligning individual and group incentives with the predominant observable features of the proposed innovation. These ideas are illustrated with three examples centered on, respectively, (a) revision of an existing data-collection instrument; (b) development of a new set of small area estimates; and (c) cost-effective acquisition and integration of alternative data (e.g., administrative or commercial records) with traditional sample survey data. Key words: Adoption and diffusion of innovations; balance of quality, risk and cost; incentives; public goods; use value and option value; utility function

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Measuring Sexual Orientation and Gender Identity (SOGI): Where We've Been, Where We're Going and New Insights (Panel)

Measuring Sexual Orientation and Gender Identity (SOGI): where we've been, where we're going, and new insights.

Eric W Jamoom *National Center for Health Statistics*

James Dahlhamer, NCHS; Eric Jamoom, NCHS

Jim Dahlhamer *National Ctr for Hlth Statistics*

Eric Jamoom *National Center for Health Statistics*

Carolina Milesi *NORC at University of Chicago*

Darby Steiger *Westat*

Sexual and gender minority (SGM) populations are becoming more visible in social and political life, yet there remains a lack of data on the characteristics and well-being of these groups. The Office of Management and Budget (OMB) convened the Federal Interagency Working Group on Measuring Sexual Orientation and Gender Identity (SOGI IWG) to begin addressing the deficiency of information for these populations and the concerns surrounding methodological issues in collecting such data. The first paper will present the findings from a meta-analysis of evaluations of SOGI measurement error in surveys to date, and identify gaps from the literature. While sexual and gender identity measurement has been used and evaluated across population surveys (e.g., NSFG, NHIS, and CHIS), evaluation is limited in the workplace. The next paper will present findings that measure SOGI in the workforce, including additional insights from using web probes for measuring respondent attitudes and understandings about these concepts. This work will inform how employers can communicate the collection of SOGI, including proactively addressing any employee concerns. Despite progress in the development and use of a measure of sexual orientation, comprehension issues have been identified – particularly among Spanish speakers. The next paper describes insights from cognitive interviews that explore the effectiveness of alternate Spanish-language response options for sexual identity to mitigate earlier identified comprehension concerns. As recommended by the SOGI IWG, further examination of SOGI concepts in adolescent populations is recommended. This last paper presents findings from a series of cognitive interviews that were conducted with youth ages 12 to 21 residing both in facilities and the general population. This presentation includes findings across two rounds of interviewing, discussing how questions were modified for the 2nd round of testing, and the results of each round. Considerations for future measurement among adolescents will also be presented.

Measuring Sexual Orientation and Gender Identity – What's been done and where are we going?

James Dahlhamer, NCHS; Eric Jamoom, NCHS

Sexual and gender minority (SGM) populations are becoming more visible in social and political life, yet there remains a lack of data on the characteristics and well-being of these groups. More representative and better quality data need to be collected. The U.S. Federal Government is taking several steps to more efficiently coordinate data collection efforts across its many departments. The Office of Management and Budget (OMB) convened the Federal Interagency Working Group on Measuring Sexual Orientation and Gender Identity (SOGI IWG) to begin addressing the deficiency of information for these populations and the concerns surrounding methodological issues in collecting such data. A meta-analysis

of evaluations of SOGI measurement error in surveys, a comprehensive literature review was conducted. The review focused primarily on SOGI measurement research from surveys primarily conducted in the United States. This included a review of Federal agency directed—or sponsored—reports, proceedings papers of the American Association for Public Opinion Research (AAPOR) conferences, and literature searches related to the topic of SOGI measurement. The scope was limited to publications occurring between January 2011 and July 2016 (exceptions as noted). This session will present information around work from this group on evaluations of SOGI measures, and identifying gaps from the literature.

Informing the measurement of sexual orientation and gender identity in the work place.

Eric Jamoom, NCHS; Paul Scanlon, NCHS

Given recent advances and interest in measuring Sexual and Gender Minority (SGM) populations, attention has expanded to the workplace. Addressing the feasibility of measuring sexual orientation and gender identity (SOGI) information like other demographic information is explored through a Federal SOGI workforce pilot that asks demographic questions, as well as a combination of open and closed-ended probes about gender identity and sexual identity. Conducted across 2 rounds of data collection through the web, respondents were asked three SOGI questions—two related to gender identity (biological sex at birth and current gender identity), and one on sexual identity. This presentation will focus on the analysis of the emerging topic of gender and sexual identity, specifically within the workforce. While these questions have been used and evaluated across population surveys like (NSFG, NHIS, and CHIS), evaluation is limited within the context of the work place. In addition to the SOGI measures, web probes were also used to measure respondent attitudes and understandings around discussing these concepts. Respondents were asked about their level of comfort discussing SOGI, and the risks and benefits to providing such information to their employers. This work will inform how employers can communicate the benefits for collecting this information, including proactively addressing any employee concerns.

Identifying Sexual Orientation among Adult Spanish Speakers

Carolina Milesi, Heather M. Morrison, Rene Bautista, and Michael J. Stern, NORC at U of Chicago

Items to accurately measure sexual orientation and gender identity in surveys are needed to assess disparities facing LGBT populations. Despite progress in the development and use of a measure of sexual orientation, comprehension issues have been identified – particularly among Spanish speakers. Stern, et.al. 2016, found that in cognitive interviews conducted with older adults, Spanish speakers exhibited comprehension problems with the response options offered; of particular concern was the term “heterosexual.” More than half of Spanish speakers were unable to select “heterosexual” during the interview even though they had stated that they were “straight” in the screener. Among these “straight” respondents who struggled to comprehend this term, about half answered “I don’t know how to answer” and the other half selected “Something else.” To explore the effectiveness of alternate Spanish-language response options in mitigating this comprehension problem, NORC, in collaboration with the Centers for Medicare & Medicaid Services (CMS) and the National Center for Health Statistics (NCHS), conducted cognitive interviews with a diverse population of Spanish-speaking adults. The cognitive research assessed comprehension and reactions to the sexual orientation measure when presented with two alternative response options. The presentation will explore findings from the 39 cognitive interviews conducted for the current research as well as compare our results to those obtained in Stern, et.al. 2016. We find that Spanish speakers were able to respond more accurately to the measure of sexual orientation when the response options were modified. Although the modified response options is

not a direct translation of the English, it is more readily understood by Spanish speakers because it more effectively conveys the intended concept.

Improving the Measurement of Sexual Orientation and Gender Identity Among Youth

Darby Steiger, Leanne Heaton, Jessica Behm and Crystal MacAllum, Westat; Jessica Stroop, BJS

Much attention has been paid in recent years to improving the accuracy of identifying and measuring sexual and gender minority (SGM) populations. In previous research, sexual orientation and gender identity (SOGI) were assessed in a single item rather than being studied as distinct constructs (e.g. sexual attraction, sexual identity, gender identity) often leading to flawed estimates (Austin et al 2007; DeMaio et al., 2013). A recommendation from the 2016 Office of Management and Budget (OMB) Federal Interagency Working Group (FIWG) on SOGI measurement called for further examination of SOGI concepts in adolescent populations. Prior cognitive testing of adolescents in the general population found the terminology used within items, and the inclusion of multiple items assessing the same concept improved the measurement of SOGI concepts (Austin et al., 2007). To support development of new items for the third wave of the National Survey of Youth in Custody (NSYC-3), a nationally representative survey of adjudicated youth in juvenile correctional facilities, a series of cognitive interviews were conducted with youth ages 12 to 21 residing both in facilities and in the general population. One of the goals of the cognitive testing was to test the SOGI items developed by the working group. A total of 20 facility interviews and 20 general population interviews were conducted in 2016. Gender identity was asked using a two-step measure comparable to that used on the National Crime Victimization Survey (sex assigned at birth, and current gender identity), and sexual orientation was asked as two questions (NCVS item on sexual orientation plus additional item on sexual attraction). This paper will present the findings from the two rounds of interviewing, discussing how questions were modified for the 2nd round of testing, and the results of the each round. Considerations for future measurement among adolescents will be presented.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Validating Health Insurance Coverage in Surveys Post-Reform (Panel)

Validating Health Insurance Coverage in Surveys Post-Reform

Joanne Pascale *US Census Bureau*

Jeanette Ziegenfuss

Kathleen Call *Univ. of Minnesota, SHADAC*

Angela Fertig *Medica Research Institute*

Don Oellerich *US Department of Health and Human Services*

Jeanette Ziegenfuss *HealthPartners Institute for Education and Research*

Brett O'Hara *US Census Bureau*

Jennifer Day *US Census Bureau*

Marina Vornovitsky *US Census Bureau*

Surveys generally derive estimates of the uninsured by asking about coverage through a range of different sources, and then identifying those with no reported source of coverage. While research suggests that reporting of individual coverage type is problematic, evidence also suggests the ultimate measure of the uninsured is reasonably accurate, possibly because the reporting error of coverage type “nets out.” For example, there is a large literature on the Medicaid undercount indicating that upwards of 35 percent of those enrolled according to the records are not reported to have Medicaid in surveys. However, the vast majority of these individuals are reported to have another type of coverage. Thus, they are not misclassified as uninsured, but their Medicaid is under-reported and other types of coverage are over-reported. The CHIME study was launched to examine this kind of error in a more comprehensive way than any prior study has, and to do so in a post-health-reform era with two major federal government surveys – the Current Population Survey Annual Social and Economic Supplement (CPS) and the American Community Survey (ACS). The study began with a sample of enrollees across a range of coverage types or strata (public, employer-sponsored, and directly-purchased both within and outside the marketplace) provided by Medica (a private health plan based in Minnesota). Enrollees were randomly assigned to either the CPS or ACS health insurance module. Content of the papers are: (1) General context for validation studies, and methods for the CHIME study; (2) How coverage type was categorized in the CPS; (3) Results of multiple accuracy metrics including an assessment of absolute accuracy (records versus survey estimate) and relative accuracy (absolute accuracy of CPS versus ACS); (4) Covariates of accurate versus inaccurate reporters; (5) Results in the context of plans for production CPS edits, processing and final estimates.

An Overview of Validation Studies on Health Insurance Reporting in Surveys and Methods Overview of the CHIME Study

Jeanette Ziegenfuss

Self-reporting in surveys is required to quantify concepts where there are no administrative records or where administrative records cannot be aggregated across institutions to produce estimates across underlying populations of interest. In the case of health insurance, there is no administrative list of the uninsured, nor a comprehensive data source to produce national estimates of public insurance coverage, necessitating the reliance on self-reported survey data. While all self-reported data are subject to error, the complexity of health insurance coverage type makes it particularly vulnerable to

measurement error. Strategies to understand the presence, magnitude and directionality of measurement error take three primary forms; experimental studies, qualitative studies and record check studies. Record check studies, where known statuses from administrative records are compared to self-reported survey data of the same, are the gold standard for understanding measurement error. In this presentation, we detail the approach of the CHIME study to understand reporting accuracy of coverage type and the uninsured in two major federal surveys – the Current Population Survey Annual Social and Economic Supplement (CPS) and the American Community Survey (ACS) – by comparing survey reports of coverage to enrollment records from a private health plan. We highlight the strengths and limitations of this approach and describe the incremental improvements of the study design compared to the limited few other studies that have employed similar approaches.

Categorizing Type of Health Insurance Coverage in the Redesigned Current Population Survey

Joanne Pascale

In many surveys, including the ACS, categorizing a respondent's source of health insurance coverage is straightforward because the questionnaire is essentially a laundry list of yes/no questions on distinct coverage types (employer-sponsored, Medicaid, etc.). The redesigned CPS questionnaire, however, is considerably more complicated and begins with a general question on any coverage at all, and then uses follow up questions about the features of the coverage (general source, government program name, marketplace, premium and subsidization) that can be used to categorize coverage type. Furthermore, the introduction of the marketplace blurred the lines between public and private coverage. For example, coverage reported to be obtained through the government, on the marketplace, with a subsidized premium could be marketplace coverage (i.e., private), or it could be CHIP, Medicaid or another public program that requires enrollees to pay a portion of the monthly premium. In addition to this inherent ambiguity of distinguishing public from private coverage, respondents themselves could provide answers that are inconsistent. For example, respondents could report that their coverage is through the government, called Medicaid, and that there is an unsubsidized premium. Thus for the redesigned CPS, judgements are necessary to determine how answers to questions about features of the coverage should best be pieced together to categorize coverage. That is, an algorithm is needed. The CHIME study provided a unique and powerful opportunity to explore alternative strategies for categorizing coverage type utilizing self-reported data since actual enrollment status across a range of coverage types was known a priori. This paper will first provide results of reporting accuracy of the individual features of coverage (e.g., general source, program name, premium, subsidization) within strata. It will then provide accuracy metrics of self-reported coverage type under several alternative algorithms and discuss strategies for maximizing accurate coverage type categorization while minimizing miscategorization.

Validating Self-Reported Health Insurance Coverage: A Comparison of Measurement Error between Administrative Records and Survey Data

Angela Fertig

National surveys that collect information on health insurance status produce different estimates of health insurance coverage. Previous research indicates that much of the variation in the estimates is rooted in subtle differences in the questionnaires. This study assessed the reporting accuracy of private and public coverage and the uninsured in two major federal surveys – the Current Population Survey Annual Social and Economic Supplement (CPS) and the American Community Survey (ACS) – by comparing survey reports of coverage to enrollment records from a private health plan. The ACS uses a simple “laundry list” question where the CPS employs a complex funneling design which probes source of coverage, plan type, as well as premium and subsidy status. We report differences in reporting

accuracy within and across these two surveys using several measures of accuracy. Our findings indicate that 1) very few individuals with known coverage were misreported as uninsured, although this varied somewhat by the type of coverage they had and by survey treatment; 2) the reporting accuracy is significantly higher in the CPS than in the ACS, although the differences are generally small in magnitude; and 3) the reporting accuracy is lowest for non-group/marketplace coverage in both surveys. However, specific estimates of non-group/marketplace coverage should be used with an understanding of the difficulties of assessing this coverage type from surveys given post-ACA changes in how insurance is accessed and structured blurring the lines between public and private coverage.

Who gets it right? Characteristics Associated with Accurate and Inaccurate Self-Reported Health Insurance Coverage

Kathleen Call

Survey reports of health insurance status (insured or not) and source of coverage has measurement error, which is not randomly distributed. For example, 95-97% of people known to have employer sponsored insurance (ESI) are reported as having ESI, 1-2% fail to report any coverage, and others mistakenly report ESI who do not appear to have this source of coverage. By contrast, 78-83% of people known to have public insurance (e.g., Medicaid) are accurately reported as having public insurance, yet 5-6% are reported as lacking insurance. Using multivariate models, we examine characteristics associated with accurate survey reports of source of insurance coverage that are confirmed in administrative records (true positives) and misreports of a lack of insurance (false negatives). Past research, limited to public insurance enrollees, demonstrates greater Medicaid reporting accuracy among respondents for child versus adult enrollees, those with low incomes, who are unemployed, have a high school degree or less, participate in other programs (e.g., TANF, SSI) and receive medical care (based on claims data). In the current study we describe correlates of accurate and inaccurate reporting for people with both private (ESI, non-group, marketplace) and public insurance. Using data available in the CHIME survey and health plan administrative records our correlates include a standard set of demographic characteristics (e.g., sex, age, race/ethnicity, marital and employment status, education, income, household size, employer size), health status (e.g., self-reported health status and a health risk measure derived from claims data), characteristics of the survey reporting task (e.g., household size, relationship between household respondent and other household members), and characteristics of the coverage (e.g., enrollee or policyholder status, premium charge, receipt of subsidy, duration, recency, and stability of coverage). We also explore differences in correlates across the two surveys in the CHIME experiment: the Current Population Survey and the American Community Survey.

The CPS Redesign in the Context of Production Estimates of Coverage: Next Steps

Brett O'Hara

After more than a decade of research and testing, the Census Bureau redesigned questions on health insurance coverage in the Current Population Survey. This new set of questions expands the scope of the health insurance coverage, improves question understanding and lessens respondent burden, and provides a strong new baseline for measuring health insurance coverage. The redesigned questions differ from the old questionnaire in three ways: reference period, coverage types, and household-level design. This question begins the conversation about health insurance and improves responses to questions about health insurance coverage in the previous calendar year. It also starts with general coverage questions and drills down to specific types of coverage via different paths depending on previous answers. From these redesigned questions, we are able to get better measures of health insurance coverage and to enhance our understanding of health insurance churn and receipt over the

course of the year. We discuss the complexity of the new survey instrument, the intricacies of health insurance measurement, and the challenge of designing the questionnaire and its new processing system.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Questionnaire Design: Response Options, Response Format and Data Quality
Comparing the Performance of Agree/Disagree and Item-Specific Questions over PCs and Smartphones

Jan Karem Höhne *University of Göttingen*

Melanie Revilla *RECSM-Universitat Pompeu Fabra*

Timo Lenzner *GESIS - Leibniz Institute for the Social Sciences*

In quantitative social research the use of agree/disagree (A/D) questions (i.e. response categories are based on an agreement continuum), is a common and very popular methodological technique to measure attitudes and opinions of respondents. For instance, this question format is frequently used in the Eurobarometer, the ANES and the ISSP. Theoretical considerations, however, suggest that A/D questions require an effortful and intricate cognitive information processing. For this reason, a variety of survey scientists recommend the use of item-specific (IS) questions (i.e. response categories address the underlying dimension of attitudes and opinions directly) since they seem to be less burdensome. In the current study, we investigate cognitive effort (by means of response times and answer changes) and response quality (by means of survey satisficing indicators) associated with A/D and IS questions over PCs and smartphones. To investigate cognitive effort and response quality, we used the Netquest opt-in access panel in Spain and applied a split-ballot design with four experimental groups defined by device type (PC vs. smartphone) and question format (A/D vs. IS) resulting in a 2-by-2 research design. The first and second group contained $n = 300$ respondents answering A/D or IS questions on PCs, respectively. The third and fourth group, in contrast, contained $n = 400$ respondents answering A/D or IS questions on smartphones, respectively. Although the data analysis is still imminent, we expect – against current theoretical considerations – to observe longer response times and more answer changes for the IS than for the A/D question format, irrespective of the device type (PCs or smartphones). In addition, we also expect to observe higher response quality for IS than A/D questions.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Mobile Measurement: Scales and Grids

What's the Best Size for Matrix-Style Questions in Online Surveys?

Rebecca Hofstein Grady *SurveyMonkey*

Mingnan Liu *SurveyMonkey*

Using matrix questions in a survey can be easier for a surveyor to set up (as opposed to a series of single item-by-item questions with identical response options), but can easily become overwhelming for participants. Across two experiments, we wanted to determine the optimal size of the grid in matrix-style questions that would best balance respondent experience and data quality. In Study 1 (N=1,700), panelists on SurveyMonkey Audience took a survey that had 20 questions presented in a grid. All participants saw the same questions, but we varied the number of columns (3, 5, or 7) and the number of rows on each page (5, 10, or 20). We examined both data quality outcomes (e.g., straight-lining, response rate, internal reliability of scales) and subjective experience outcomes (e.g., rating of survey layout, completion time). For row size, respondents were significantly more likely to drop out and rate the survey as more difficult as the row size increased, even though that meant they had fewer pages overall. For column size, 7 columns increased the completion time of the survey, while 3 columns produced lower reliability coefficients. The best overall size was a 5x5 matrix. However, the 20 questions were made of up 4 separate scales which did not cross page breaks. In Study 2 (N=1,250), we tested to see if having fewer rows on a page was still advantageous when using 20-item scales that crossed page breaks. We found that participant experiences were still best in the 5-row condition, and scale reliability did not decrease even though the page breaks meant the scales were not all on one page. These results suggest that having 5 or fewer rows/items per page, and 4-5 columns/response options, gives the optimal experience when using matrix-style questions.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Improving Data Quality of Health Surveys

The Role of Measurement Error Due to Old Age: Determining the Amount of Discrepancies in Pension Reports Using Administrative Data from Share-RV

Patrick Lazarevič *TU Dortmund University*

A basic assumption for all comparative analyses is that measurements are comparable between cases. This premise is violated if groups of observations tend to have higher measurement errors than other groups. Since the functioning of a person's memory declines with age, one can expect that, on average, older respondents have greater difficulties answering factual questions in an interview. These greater difficulties might potentially result in two outcomes: Firstly, older respondents might have greater difficulties reporting the correct answer and therefore rather give a rough estimate based on their best knowledge. In this case, respondents of greater age should display larger measurement errors than younger respondents. Secondly, older respondents, might refuse the answer altogether in order to not disclose their uncertainty or avoid misreporting. In this case, the consequence would be item-nonresponse. To determine the role of age-specific measurement errors, administrative data of the German Pension Fund, as a reference value, were linked to 1.182 respondents' self-reports of pensions. We compared measurement errors of six age-groups and used logistic and linear regression analyses to explain the probability of sizable measurement errors and their amount as well as the probability of item-nonresponse. Regarding descriptive statistics, an almost consistent growth of discrepancies with increasing age is evident. This applies also to the share of above-average (greater than the 50th percentile), high (greater than the 75th percentile), and extreme (greater than the 90th percentile) discrepancies between self-reports and administrative data per age group. As a result of multivariate analyses, the occurrence of sizable discrepancies and item-nonresponse as well as the amount of discrepancies are, among other factors, related to higher age. These results have implications for survey methodology, as they draw attention to a source of measurement errors. They highlight the necessity to consider the role of age when measuring factual information.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Mobile Measurement: Scales and Grids

How to Ask About Mobile Device Usage Behavior? Qualitative and Quantitative Research on Survey Scales.

Tim Oltman *Nielsen*

Lauren Walton *Nielsen*

Jennifer Hunsecker *Nielsen - Oldsmar, FL*

According to Pew Research Center, 68% of Americans have smartphones, 45% have tablet computers and ownership is on the rise year over year. Understanding how these devices are used and shared among household members is of great interest to media and marketing companies as well as academic researchers. This Nielsen research presents findings on how to ask about how people use their mobile devices from a quantitative and qualitative perspective. In November 2015 we fielded a study using GfK's Knowledge Panel to ask respondents about the frequency that their household members used mobile devices. It also asked respondents what percentage of the time these same household members used the devices. We find that the 4 category behavioral frequency question was related to the percentage of the people used the device, but there was a lot of variation within the categories. Among adults reported as using a tablet "at least once a day," we found they were using that device on average 49% of the time. In comparison, adults that were reported to be using the "less often" than once a week were using the device about 12% of the time. During July 2016, qualitative research was conducted with 16 cognitive interviews. Participants were presented with a computerized mock-up of a survey that asked questions about how often they used smartphones and tablets alone and in conjunction with others. About half of the respondents did not understand questions that asked them indicate the percentage of time household members used devices to consume different categories of media content. Additionally, respondents struggled with answering questions about co-viewing, either due to lack of understanding the question or because their mental model of shared media consumption did not align with the questions asked.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Mobile Measurement: Scales and Grids

Grading the Grids: What Works and What Doesn't

Mario Callegaro *Google*

Yongwei Yang *Google*

Natalie Rojowsky-Kessel *Google*

Marni Hirschorn *Ipsos*

Amy Hill *Ipsos*

Cecile Carre *Ipsos*

Grids (or matrix, table) are commonly used on self-administered surveys. In order to optimize online surveys for smartphones, grid designs aiming for small-screen devices are emerging. In this study we investigate four research questions regarding the effectiveness and drawbacks of different grid designs, more specifically do the grid design effect: Data quality, as indicated by breakoffs, satisfying behaviors and response errors? Response time? Response distributions? Inter-relationships among questions? We conducted an experiment in August 2016 testing grid designs on three types of answer scales: a 7-point fully-labeled rating scale, a 5-point fully-labeled rating scale, and a 6-point fully-labeled frequency scale. Respondents from the US and Japan to an online survey were randomly assigned to one of three conditions: (a) no grid, where each question was presented on a separate screen; (b) responsive grid, where a grid is shown on large screens and as single-column vertical table on small screens (with question stem fixed as header); (c) progressive grid, where grouped questions were presented screen-by-screen with question stem and sub-questions (stubs) fixed on top. Quotas were enforced so that half of the respondents completed the survey on large-screen devices (desktop/tablet computers) and the other half on smartphones. Respondents were 600 per grid condition per screen size per country. Initial findings showed that progressive grid had less straightlining and response errors whereas responsive grid had less break-offs. Differences were also found between grid designs in terms of response time and response distributions; however patterns varied by country, screen size and answer scales. Further analysis will explore the effect of grid design on question inter-relationships. While visual and interactive features impact the utility of grid designs, we found that the effects might vary by question types, screen sizes, and countries. More experiments are needed to explore designs truly optimized for online surveys.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Improving Data Quality of Health Surveys

Health Insurance Statement Usage and Respondent Characteristics

Jennifer Vanicek *NORC at the University of Chicago*

Lauren McNamara *NORC at the University of Chicago*

Nicholas Schluterman *Centers for Medicare and Medicaid Services*

The Medicare Current Beneficiary Survey (MCBS) is a primary source of data on health care costs for the Medicare population. The MCBS is a continuous, multipurpose survey of a nationally representative sample of the Medicare population, conducted by the Centers for Medicare & Medicaid Services (CMS) through a contract with NORC at the University of Chicago. The MCBS collects detailed data about health care utilization and associated costs, which depend heavily on accurate recall and documentation by the respondents. Respondents are encouraged to retain their health insurance statements to assist with data collection of health care costs, and approximately 60% of respondents reporting events present at least one health insurance statement such as a Medicare Summary Notice or Explanation of Benefits statement. When respondents have statements corresponding to survey-reported events, more accurate data on costs and payments, particularly payments by non-Medicare sources, can be collected. This paper will use data collected in fall 2015 and winter 2016 to explore the characteristics of MCBS respondents who tend to retain and provide insurance statements for use in the interview. In particular, are there certain types of respondents who are more likely to save and present health insurance statements during the interview? Does this vary by factors such as respondent demographics, proxy interview, type of insurance coverage, or health status? The paper will also discuss current protocols for encouraging respondents to retain their health records. Knowing more about who is likely to retain and present statements during the interview can help inform these training and field protocols, and increase the number of respondents who keep these documents, thereby enhancing the quality of cost data collected.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Questionnaire Design: Response Options, Response Format and Data Quality
Reducing Measurement Error in Interviewer-administered Surveys - The Effects of Response Scale Format and Survey Mode on Sensitive Attitudinal Questions
Chariklia Hoefig Bw Center for Military History and Social Science

The presentation examines and discusses scale format and mode effects on data quality in surveys about political attitudes towards security and defense policy, which are assumed to be sensitive topics. In order to examine these measurement effects, a large-scale method experiment was conducted in two concurrent surveys (CATI and CAPI) with probability samples of the German population. Both surveys included the same 44 items with varying content, stimuli, response labels and number of response options (split-ballot: 4-, 5- and 7-point scale). Hence, the robustness of the effects can be assessed. To analyze interaction effects of the response scale and the mode with interviewees, various characteristics were documented. To get more information about the interviewee and the interview situation, the interviewer was asked to assess respondents' characteristics and behavior during the interview and the interview situation in general. In addition, metadata like response latency and time of day were recorded. In order to control for interaction with interviewer characteristics, socio-demographics and job experience of the interviewers were included in the dataset. The unique setting of the large-scale experiment allows disentangling measurement effects due to the instrument (scale format and survey mode), the interviewer, the respondent and the situation on different aspects of data quality. Findings and implications for survey design will be discussed.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Improving Data Quality of Health Surveys

Validating Self-Reported Incidence of Two Chronic Conditions among Ohio's Newly Enrolled Medicaid Population with Administrative Claims Data

Rachel Tumin *Government Resource Center*
Michelle Menegay *Government Resource Center*
Michael Nau *Government Resource Center*
Daniel Weston *Government Resource Center*
Marcus Berzofsky *RTI International*
Timothy R Sahr *Government Resource Center*

Beginning January 1, 2014, most Ohioans age 19-64 years with incomes $\leq 138\%$ of the federal poverty level became eligible for Medicaid benefits as part of the Affordable Care Act. The Ohio Medicaid Group VIII Assessment examined the impact of Ohio's Medicaid expansion for these newly eligible enrollees using a mixed data collection approach, which included a stratified random telephone survey and Medicaid administrative data. These survey and administrative data were linked together because the survey predominantly collected data not captured by Medicaid claims. However, both data sources provided estimates of the incidence of two chronic conditions among newly eligible enrollees for the same time period since gaining coverage, enabling us to evaluate whether the survey and administrative data generated similar estimates. The claims yielded the percentages of enrollees with a diagnosis code for high cholesterol or diabetes, and the survey provided the percentages of enrollees reporting being first told they had either condition since getting Medicaid coverage. Study data have only recently become available and preliminary analysis showed that these data sources yielded substantially different estimates: from the claims analysis, 21.0% had a high cholesterol diagnosis and 11.5% had a diabetes diagnosis; the corresponding weighted survey estimates were 12.4% and 3.9%, respectively. There are likely several factors contributing to this measurement error, and future analyses will explore the implications of using different coding definitions for the claims-based estimates, examine differences in these population-level incidence estimates by sociodemographic characteristics, and conduct a sensitivity/specificity analysis using respondents' linked administrative claims and survey data. These results will provide insight into the potential strengths and limitations of relying solely on self-reported or administrative data to measure chronic disease incidence in a low-income population. Future research should include exploring how to integrate these complementary data sources into policy considerations.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Improving Data Quality of Health Surveys

Effects of Survey Mode on Responses to the CAHPS Hospice Survey of Care Experience: Results from a Randomized Experiment

Layla Parast *RAND Corp.*
Layla Parast *RAND Corp.*
Marc Elliott *RAND Corp.*
Katrín Hambarsoomian *RAND Corp.*
Melissa Bradley *The RAND Corp.*
Joan Teno *University of Washington*
Rebecca Anhang Price *RAND Corp.*

National implementation of the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Hospice Survey, which asks family caregivers about their recently-deceased family members' experiences with hospice care, began in 2015. Each hospice is allowed to choose from one of three modes of survey administration: Mail Only, Telephone Only, or Mixed Mode (mail with telephone follow-up). Previous studies of hospital care have found more positive evaluations of health care by telephone than by mail survey, which has required adjustment of scores based in randomized experiments to ensure comparability of results. We used a randomized experiment to compare response rates and mean responses on patient experience items by survey mode, and to determine whether survey mode adjustments were needed to fairly compare CAHPS Hospice Survey scores across hospices using different modes. We sampled 17,121 primary caregivers of patients who died while receiving hospice care from May through September 2015 from 59 large hospice programs. Within each hospice, we randomized one-third of cases to each of the three modes of data collection. One vendor collected survey data for all hospices using standardized data collection protocols. Linear and logistic regression predicted response rates and responses to 24 items measuring hospice quality. The overall response rate to the survey was 45% (AAPOR RR1)-53% Mixed Mode, 43% Mail Only, and 38% Telephone Only. There were significant mode effects for 10 of the 24 questions that compose the quality measures. Unlike results observed in the Hospital CAHPS mode experiment, hospice primary caregivers tend to respond more negatively by telephone than by mail. This unusual result may be specific to response by caregivers (rather than patients themselves) or to the nature of reporting on hospice care. Mode effects were large enough to substantially bias comparisons among hospices administering the survey in different modes unless mode adjustments are applied.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

The 2016 Election: Explanations and Implications

Defining the Working Class

Ashley Kirzinger *Kaiser Family Foundation*

Elizabeth Hamel *Kaiser Family Foundation*

During and in the wake of the 2016 presidential elections, there have been many examinations of the economic and social underpinnings of white working-class voters. The importance of group identity in American politics did not mysteriously appear with the nomination of Republican presidential candidate Donald Trump. In fact, public opinion researchers have long understood and documented the importance of group identity, and in-group/out-group politics, in attitudes as well as voting behaviors. A lot of this research has focused specifically on White identity as well as working-class identity – largely due to the decline in power of these groups in American politics over the past several decades. Our analysis draws on the work of other researchers by examining the methodological decisions that researchers make when they define “working-class” and how three common definitions can affect attitudes towards the economy, government, racial and ethnic diversity, and immigration. By choosing different variables (including education, income, job type) to define their “working class,” unsurprisingly, researchers alter the partisan, age, and gender make-up of the group under study – especially among non-Hispanic whites. Among whites, an income-based classification leads to a less Republican-leaning group while employment-type and education-based classifications lead to more Republican-leaning groups. However, the various definitions have a more diminished effect on attitudes towards economy, trade, immigration, and the future of the country. Our research demonstrates that any major differences between the various ways researchers define this group dissipate when examining attitudinal measures, thus, perhaps demonstrating that the fundamentals defining this group transcend just partisan attitudes.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

The 2016 Election: Explanations and Implications

The Effect of the 2016 Presidential Election on Consumer Optimism

Jennifer M Bouterse *Simmons Research*

Steven Millman *Simmons Research*

Given the acrimony and deep partisan divisions that defined the 2016 US Presidential election, as well as the palpable anger in the aftermath, how will Trump's and Clinton's supporters change their outlook on the economy and their own financial health? Will these differences be of a larger scale than in the 2012 election? To test these hypotheses, we will compare consumer sentiment and life events planning before and after the election for respondents of the Simmons National Consumer Survey and look for differences among self-identified Republicans, Democrats, and Independents. We will also look at the differences across ideology for conservatives, liberals and moderates. Some of the variables we will compare will include current financial outlook; future financial outlook; plans to enroll or return to college in the next 12 months; plans to change jobs in the next 12 months; plans to buy first home, new home or refinance a home in the next 12 months; and plans to retire in the next 12 months. We will then go back to the 2012 election, and once again compare consumer sentiments and life events planning before and after the election for respondents of the Simmons National Consumer Survey to see if there are different trends for the winners and losers of the election in 2012 than we see in 2016. The National Consumer Survey is a nationally representative, probabilistic paper survey produced by Simmons Research.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Mobile Measurement: Scales and Grids

The Effects on Data Quality and Response Distributions of Horizontal and Vertical Question Orientation for Different Devices in Online Surveys

Johan Martinsson *University of Gothenburg*

Elias Markstedt *University of Gothenburg*

Delia Dumitrescu *University of Gothenburg*

The increasingly mixed technical devices respondents are using to answer online surveys have caused some concern among survey researchers. Despite many efforts to adjust online surveys to smartphones our knowledge is still limited when it comes to how different response scales work on different devices. In particular, responsive design (without a complete mobile first approach) can now accentuate the differences in how a questionnaire is presented on different devices. The main concern is that respondents who use a specific device, e.g. a tablet or a smartphone, might not treat the response scales in the same way as PC or laptop users. This paper analyzes response option order effects and the large sample sizes used enables us to analyze to what extent these effects are conditional on scale length, layout direction (horizontal vs vertical) and the type of response device participants are using. This paper reports on a set of experiments where not only response option order and scale visual orientation (horizontal or vertical) is varied randomly, but also scale length. A total of four experiments are reported. The results indicate that vertically oriented response scales suffer more from primacy effects than horizontally oriented response scales, and that this is even more so for smartphone users. Further, scale length also makes a difference in that primacy effects are substantially larger for smartphone users when scales are longer than 5 scale points.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

The 2016 Election: Explanations and Implications

Understanding the 2016 U.S. Presidential Election Polls: Was Non-Response Bias a Problem?

Peter K. Enns *Cornell University*

Jonathon Schuldt *Cornell University*

Although the 2016 U.S. presidential election polls forecasted the popular vote quite accurately, public opinion surveys (especially state polls) have been heavily criticized for failing to predict Donald Trump's victory, raising questions about whether the polls suffered from non-response bias or whether the samples were representative but influenced by some other source of error. We utilize data from nearly identical surveys fielded in October 2016—a nationally representative internet sample (n=1,461) conducted by GfK and a nationally representative telephone sample (cell and landline, n=625) conducted by Cornell's Survey Research Institute—to show that survey non-response bias does not appear to explain the polling errors. We rely on a survey question that allows us to identify those who leaned toward Clinton or Trump but did not express a direct vote intention: "If you HAD to choose, which presidential candidate do you find to be more truthful?" Across both surveys, almost 97 percent of respondents who expressed a vote intention for either candidate indicated that candidate as more truthful, suggesting this "truthful" question is an excellent proxy for vote choice. Importantly, almost all respondents answered the "truthful" question, allowing us to estimate candidate choice among those who indicated they were undecided or not planning to vote. If non-response bias explained the forecasting error, analyzing a different question should not improve the accuracy of our estimates. However, not only does this "truthful" question forecast the final vote share within two percentage points in both surveys, but a Multi-Level Regression and Post-Stratification (MRP) analysis to generate state-level vote-share estimates revealed Trump winning multiple key swing states that election polls failed to predict. We believe these findings offer a powerful affirmation of polling methodology and a path to avoiding similar forecasting errors in the future.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Evaluating Election Polling Accuracy

Error estimation in election polls

Ronaldo Iachan *ICF International*

John Boyle *ICF International*

Election polls currently adopt a range of different methodologies including dual frame random digit (RDD) telephone surveys, internet panel surveys, and new combinations of these approaches. The estimation and reporting of survey error is typically encapsulated in a margin of error (MOE) estimate that understates the actual errors in a multitude of ways. This paper explores these sources of error and how they may impact the overall estimates. While some of these sources are unique to, or at least more egregious in, election polls, they are also observed in surveys in general. The polling estimation is a two-stage process which includes a model for estimating likely voters and then estimating voters for a candidate among likely voters. The sampling variance should include the two components of the variance which are described in the paper. In addition to sampling error, the total error includes non-coverage biases, non-response biases and measurement error. We also discuss briefly how these errors may have induced larger and systematic biases in the 2016 presidential elections, as well as their differential magnitude by data collection methodology. We discuss how some weighting methodologies correct for some but far from all of these imbalances. We conclude by showing how the MOE may be a crude and misleading representation of accuracy in most state and national polls.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Evaluating Election Polling Accuracy

Predicting 2016 State Election Results with a National Tracking Poll and MRP

Chad Kiewiet de Jonge *Langer Research Associates*

Gary Langer *Langer Research Associates*

This paper presents state-level estimates of the 2016 presidential election using data from the ABC News/Washington Post tracking poll, which included over 7,000 likely voters interviewed during the 19 days preceding the 2016 election. To do so, the analysis employed multilevel regression with poststratification (MRP) to estimate who was likely to vote and for whom, using basic demographic variables at the individual level (gender, age, education and race) and a few state-level variables (past party vote shares and shares of each state that are black, Hispanic and evangelical white Protestants). For poststratification, the analysis used the most recent 5-year ACS estimates (2010-2014). The paper presents not only the state-level turnout and vote choice estimates from this approach, but also a variety of subgroup-level estimates, and discusses the implications for understanding the contours of the election. Overall, the MRP estimates proved highly accurate in predicting the winner in each state (missing just one) and was more precise in estimating the national popular vote than many published results. We find similar results using ABC or ABC/Post pre-election tracking data from 2004, 2008 and 2012 alike. While far from perfect in estimating the exact popular vote shares and margins in each state, the MRP approach provided a much more accurate representation of the election outcome than prevailing polling averages and forecasts. The paper concludes by discussing how pollsters and others – to the extent they are focused on predicting election outcomes – could make use of MRP as an alternative approach to weighting and likely voter modeling.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

The 2016 Election: Explanations and Implications

The 2016 Election: How and Why it's President Trump

Gary Langer *Langer Research Associates*

Gregory Holyk *Langer Research Associates*

Chad Kiewiet De Jonge *Langer Research Associates*

Sofi Sinozich *Langer Research Associates*

Controversy over predictions of the outcome of the 2016 presidential election threatens to obscure the true value and higher purpose of pre-election polling: to understand how and why the public comes to its choices. We produced 18 full-length national public opinion polls for ABC News in the 2016 contest, culminating in a 19-day pre-election tracking poll, and analyzed exit poll results in party primaries and caucuses in 25 states as well as the national and state general election exit polls. Our presentation will review and synthesize results of this intensive research effort to explain the election of Donald Trump as the nation's 45th president.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Evaluating Election Polling Accuracy

Assessing the Accuracy of Pre-Election Polls: 2008 - 2012

Jon A Krosnick *Stanford University*

Jon Krosnick *Stanford University*

Ahra Cho *Stanford University*

Amanda McLean *Stanford University*

Christopher Middleton *Stanford University*

David Kay *Stanford University*

Joseph Abruzzo *Stanford University*

Jelani Munroe *Stanford University*

Mark Carrington *Stanford University*

Research on the accuracy of pre-election polls has identified a wide array of poll characteristics that could affect the accuracy of polls. Drawing from the literature on survey accuracy, in general, and the literature on the accuracy of pre-election polls, in particular, this study examines various hypotheses regarding the relationships among the characteristics and accuracy of pre-election polls. Based on a dataset consisting 2,688 pre-election polls for the presidential, senatorial, congressional, and gubernatorial elections during the 2008 - 2012 time period, this study tests these hypotheses using three different measures of poll accuracy: the average absolute error, the absolute margin of victory error, and the partisan error (a measure of the degree to which the poll result is biased in favor of one of the parties). This study takes into account the four different roles polling firms, media organizations, and academic institutions may play in conducting polls, namely, designer, collector, sponsor, and publisher. A systematic, case-by-case data collection method was employed to identify the roles these firms and organizations played in conducting polls. Moreover, for each of the four roles a firm or organization may play in a given poll, this study collects partisanship data to provide an in-depth understanding of the partisan characteristic of individual polls. In addition to testing the existing theories of poll accuracy on an extensive scope, this study provides an assessment of the impact of polling firm partisanship on the partisan error of the polls. Initial analyses of the dataset lend support to several hypotheses proposed by previous studies and suggest that days in the field, sample size, office of the election, mode of the poll, and partisanship characteristic of the poll are significant determinants of the different measures of poll accuracy.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Mobile Measurement: Scales and Grids

Expanding Alternatives: The Accordion Grid as an Alternative to the Traditional Grid Format

Nicole R Buttermore *GfK*
Frances MBarlas *GfK*
Randall K Thomas *GfK*

Grid – or matrix format – questions are used routinely in online surveys when researchers want to assess multiple items using the same response format. However, grids can be problematic with smaller screen sizes such as those of smartphones. Horizontal (left-right) scrolling is often required to view all response options, making it impossible to view both the item text and all response options at the same time and raising the possibility that respondents will be less likely to select response options not visible on the screen. In the accordion grid format, respondents see the item text listed vertically, then can click on each item to reveal the response scale in a standard single response format. Using this approach, respondents can see the entire response scale for each item, along with the response items on a single screen. However, vertical response presentation has been associated with an increased likelihood of response order effects. We conducted several experiments using sample from KnowledgePanel®, GfK's probability-based online panel, in which respondents were randomly assigned to complete either an accordion or traditional grid. The time it took for respondents to complete each grid type was comparable, and we found few differences in results across the two types of grid for both dichotomous grids and unipolar scales with three or more response options. We found similar concurrent validity for both grid types when correlating the results with behavioral data and little evidence of response order effects. Respondents provided similar ratings of the ease and accuracy with which they could respond to the traditional and accordion grid scales. Taken together, the results suggest that the accordion grid is a promising alternative to the traditional grid design, especially given the increase in the proportion of online survey respondents who complete surveys using smartphones and other mobile devices.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

The 2016 Election: Explanations and Implications

Trump victory in the Great Lakes – a vote to change the status quo?

Pablo Diego-Rosell *The Gallup Organization*

Diana Liu *The Gallup Organization*

Jonathan Rothwell *The Gallup Organization*

The outcome of the 2016 presidential election represented a surprise for most pollsters, with the Great Lakes region as the epicenter of the upset. Trump victories in the region were as decisive as they were unexpected – Pennsylvania and Michigan last voted for a republican presidential candidate in 1988 and Wisconsin in 1984. Post-electoral analysis showed low turnout amongst Clinton's base, along with a defection of blue-collar and middle-class white voters to Trump, and an above average performance of Trump among rural voters. It was hypothesized at the time that the Democratic candidate had failed to connect with these voters' economic concerns. After years of deindustrialization and stagnating living standards, these voters were driven to the Trump camp more by a desire to change the status quo than by a favourable opinion of the Republican candidate. We examine this hypothesis by looking at well-being trends in the region using Gallup Daily tracking data. The Gallup Daily poll is a daily survey asking U.S. adults about their views on the most important issues affecting the United States and the world. The poll is administered 350 days a year, with a daily sample ranging from 500 to 1,000 U.S. adults. This study includes a total of 2.4 million interviews collected between 2008 and 2016. Our results show that well-being trends in the Great Lakes region were similar to other parts of the country. Differential trends in a given region did not predict the change in GOP vote, but average subjective well-being (SWB) levels did. Moreover, average SWB levels were better predictors of the Trump surge at the county level than standard economic indicators such as income or unemployment. We finally explore the relative contribution of different sub-domains of SWB, with a particular focus on perceived health, community well-being and workplace engagement.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Improving Data Quality of Health Surveys

What Worked and What Didn't; Changes for the 2017 National Survey of Children's Health

Jason Fields *U.S. Census Bureau - Washington, DC*

Jason Fields *U.S. Census Bureau - Washington, DC*

Reem Ghandour *HHS/HRSA/MCHB/OER/DE*

Jessica Jones *HHS/HRSA/MCHB/OER/DE*

Leah Meyer *U.S. Census Bureau*

The Maternal and Child Health Bureau (MCHB), within the Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services (HHS) has undertaken an important transition in the National Survey of Children's Health (NSCH). Working with the Census Bureau, MCHB has transitioned the NSCH to an annual survey with a household address-based sample. The sampling approach requires identifying households with children and sampling a child from within those households to provide state-level representation of children with and without special health care needs. The 2016 NSCH was implemented based on the best available information of the mix of mode and contact approaches for the self-administered web and paper designs. The design for 2016 included a test of initial unconditional incentives and focused on a web-push strategy to reach the most households within the available budget. The 2016 NSCH experienced response rates that have been lower than expected; several factors are hypothesized to have impacted response, but remain under investigation. Given the successes of the administrative flags for sampling households with children, and the new information available from the 2016 contact and response information, we are able to make several modifications to the planning for the 2017 NSCH. For the 2017 NSCH, the contact strategy will be modified to better target households that are more likely to respond via paper questionnaires than web-based questionnaires. The contact attempts will be more frequent during the initial phases and utilize pressure sensitive letters to send reminders following the initial mailing. Subsequent mailings will be adjusted based on the analysis of the 2016 NSCH response patterns to improve mailing effectiveness. Additionally, the sampling strategy is being reviewed to more effectively identify households with children in the strata where administrative flags do not indicate children.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Questionnaire Design: Response Options, Response Format and Data Quality
Are Self-Description Scales Better than Agree/Disagree Scales in Mail and Telephone Surveys?

Jerry Timbrook *University of Nebraska-Lincoln*
Jolene Smyth *Univ. Nebraska - Lincoln*
Kristen Olson *University of Nebraska - Lincoln*

Agree/disagree scales have been criticized for being cognitively complex (Fowler, 1995), and prone to straightlining and high acquiescence (Saris et al., 2010). Construct-specific questions (tailoring the stem and response options to each item) have demonstrated superior data quality outcomes over agree/disagree scales (Dykema et al., 2012), but require substantial development time (Wall et al., 2015). Self-description scales (asking respondents how well a statement describes them from Completely to Not at all) may be a useful middle ground: they do not need to be tailored to a specific construct, and by omitting the word “agree” they may avoid tendencies toward acquiescence. However the data quality implications of self-description scales are relatively uninvestigated in self-administered or interviewer-administered surveys. In this paper, we compare data quality outcomes for an agree/disagree scale and a self-description scale on a grid of 6 positive valence questions (ex: “I consider myself a good person”) and a grid of 5 negative valence questions (“I lose my temper pretty easily”) in a mail survey (National Health, Wellbeing, and Perspectives Study (NHWPS); n=1,002, AAPOR RR1=17.8%) and 4 questions in a telephone survey (Work and Leisure Today II Survey; n=911, AAPOR RR3=7.8%). Preliminary analyses of the NHWPS indicate that for 9 of the 11 questions, more respondents selected the two most positive response options using the agree/disagree scale ($p < .05$). The agree/disagree scale also has significantly more straightlining than the self-description scale for both sets of questions (positive: $p = .047$; negative: $p = .038$). Reliability for positive questions is equal for the agree/disagree ($\alpha = .81$) and self-description scales ($\alpha = .81$; $W = 1.0$, $p = .50$), but is significantly lower on the self-description scale for negative questions ($\alpha = .76$ vs. $\alpha = .82$; $W = .76$, $p < .001$). We will also examine how this affects measurement methods using Structural Equation Modeling. This paper will conclude with implications for questionnaire design in both telephone and mail surveys.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Evaluating Election Polling Accuracy

It's Lonely at the Top: Seeking Lessons from Down-Ballot Pre-Election Polling

Andrew Baumann *Global Strategy Group*

Andrew Baumann *Global Strategy Group*

Michael Smith *Global Strategy Group*

Damla Ergun *Global Strategy Group*

The failure of polling to accurately predict the results of the 2016 Presidential election, particularly in states with higher proportions of whites without college degrees, has been the cause of great consternation. Down-ballot races, however, have received much less attention in the immediate aftermath of the election. In this presentation we use polls conducted by Global Strategy Group in down-ballot races to take stock of the predictive accuracy of these polls and to explore whether the errors in prediction in down-ballot polling follow similar or different patterns than the ones observed in the Presidential race. We include races for the U.S. Senate and House as well as gubernatorial and down-ballot state offices, including in many cases data collected in states where polling was particularly inaccurate in the Presidential race, like Pennsylvania, Indiana, and Missouri. Our early analysis of the data indicates that in some cases, polling error at the top of the ticket seemed to trickle down to down-ballot race, but in some cases, estimates for down-ballot races tended to be considerably more accurate. In this presentation, we will compare results across and within states on the presidential vs. down-ballot races with a goal to highlight the nature, direction and extent of polling errors, or lack thereof – and their causes.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

**Questionnaire Design: Response Options, Response Format and Data Quality
To Smiley Or Not To Smiley: Considerations and Experimentation to Optimize Data Quality and User
Experience for Contextual Product Satisfaction Measurement**

Aaron Sedley *Google*

Yongwei Yang *Google*

Hilary Hutchinson *Google*

Happiness Tracking Surveys (HaTS) at Google are designed to measure satisfaction with a product or feature in context of actual usage. Smiley faces have been added to a fully-labeled satisfaction scale, to increase discoverability of the survey and response rates. Sensitive to the potential variety of effects from images and visual presentation in online surveys (Tourangeau, Conrad & Couper, 2013), this presentation will describe research designed to inform and optimize Google's use of smileys in Happiness Tracking Surveys across products and platforms: 1) We explore construct alignment by capturing users' interpretations of the various smiley faces, via open-ended responses. This data shows meaningful variation across potential smiley images, which informed design decisions. 2) We assess scaling properties of smileys by measuring each smiley independently on a 0-100 scale, to calculate semantic distance between smileys in order to achieve equally-spaced intervals between scale points (Klockars & Yamagishi, 1988). 3) We describe considerations and evaluative metrics for a smiley-based scale with endpoint text labels, to be used with mobile apps and devices.

May 20th, 2017
3:30 PM - 5:00 PM
Concurrent Session I

Evaluating Election Polling Accuracy

Accuracy of National and State Polls in the 2016 Election

Amanda McLean *Stanford University*

Jon AKrosnick *Stanford University*

In the aftermath of the 2016 elections, the national popular vote polls done with random sampling vis RDD to landlines and cell phones correctly predicted the winner and yielded results close to the actual vote shares, with a small, systematic bias in favor of Secretary Clinton. But many state-level polls performed less well, often predicting the wrong winner and manifesting more error. This paper involves an analysis of the national and state polls comparing accuracy calculated in various ways and exploring how different ways of handling “don’t know” responses affect accuracy. An analysis will be reported comparing the accuracy of RDD landline/cellphone surveys vs. IVR surveys vs. opt-in Internet surveys vs. telephone surveys done with quotas in terms of a range of dependent variables.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Media, Emotion and Measurement: Understanding Drivers in Discourse and Opinion on Abortion (Panel)

Media, Emotion, and Measurement: Understanding Drivers in Discourse and Opinion on Abortion

Jill Mizell Center for Reproductive Rights

Kate Stewart and Jill Mizell

Given the heightened level of debate around abortion policy, the dramatic rise in state laws on abortion access, threats to defund Planned Parenthood, and a federal government that is hostile toward women's legal right to abortion, it is time we take a deeper look at the public's nuanced attitudes on abortion. This panel will examine how we are measuring attitudes on abortion, the interplay of abortion attitudes and discourse in social and traditional media, and the roles of emotion and information on policy preferences.

Measuring Emotion: Reactions to Different Possible Outcomes in *Whole Woman's Health v Hellerstedt*

Kate Stewart and Jill Mizell

As the electorate becomes increasingly polarized along ideological lines, it's important to understand the relationship between emotions and policy preferences, especially on traditionally "hot button" social issues. By applying an adjusted version of Plutchik's emotion classification model to various issues related to abortion, including different possible decision scenarios in a landmark Supreme Court case, the authors were able to identify emotional drivers in public opinion, and the impact of information on emotions and attitudes.

Media Coverage and Abortion Stigma

Steph Herold and Lauren Himiak

We know that public opinion is influenced by media coverage, especially on culture war issues like abortion. Yet there is little rigorous data on media coverage of abortion, including measuring if that coverage is accurate or contributes to stereotypes and misinformation about abortion. The Sea Change Program conducted two studies to investigate how abortion stigma manifests in news media. We partnered with Berkeley Media Studies Group (BMSG) to conduct a quantitative, retrospective study of news coverage of abortion in the United States. We identified trends in depictions of abortion that could reinforce stigma, such as the language used, the speakers cited, and the data that is included or left out of news narratives. Concurrently, Sea Change conducted a qualitative study to investigate how and why journalists cover abortion, including how they understand their role in covering this issue and what might make abortion different from other issues journalists report on. Our research offers recommendations for journalists and abortion rights advocates on how to improve news coverage of abortion.

Crucible of conflict: Twitter and TEXAS abortion opinions

Amanda Stevenson

Taking the debate and passage of Texas H.B. 2 as a case study, this paper examines the role of emotions in shaping individuals' participation in the resistance to the bill, as well as their framing of the bill. Drawing on a massive social media dataset, I find that emotions expressed within relationships between participants contributed to ongoing engagement. Focusing on interactions between movement elites and activists, I find that emotions contributed to activists' successful transmission of a radical framing to elites.

Emotional Response to State Abortion Restrictions: Findings from Qualitative and Quantitative Research

Tresa Udem

Since 2011, 343 anti-abortion restrictions have passed in 32 states across the country. The vast majority of voters are not aware of laws on abortion. What happens when they learn about these restrictions? A surprisingly strong emotional response. Tresa will present qualitative data showing how people process and reconcile information about new restrictions. She will also present survey data that gauges voters' initial emotional response to abortion restrictions.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Cultural Competence in Field Data Collection (Panel)

Cultural Competence in Field Data Collection

Alisu Schoua-Glusberg *Research Support Services Inc.*

Alisú Schoua-Glusberg

Alisu Schoua-Glusberg *Research Support Services Inc.*

Patricia Goerman *U.S. Census Bureau*

Leenisha Marks *RTI International*

Maichou Lor *University of Wisconsin-Madison*

Charles Lau *RTI International*

Yazmin Argen Garcia Trejo *US Census Bureau*

Melissa Baker *Kantar*

Clark Letterman *RTI International - RTP, NC*

Papers in this panel will address the common theme of Cultural Competence in Field Data Collection. The presenters will discuss cross-cultural adaptations needed in fieldwork and the importance of acknowledging and incorporating the study participants' cultural context in research plans. This in turn can contribute to increased data quality, higher response rates and better relations with study populations. Alisú Schoua-Glusberg will present an overview on the topic of cultural competence in field data collection. She will address ways to incorporate this theme throughout the survey lifecycle starting with initial instrument design through completion of data collection. The remaining presenters will focus on interactions in the field, development and tailoring of verbal interviewer messages and interviewer training to presence of third parties in the field and whether they may hinder or help with the interview process. Kathleen Kephart's presentation will examine current interviewer training practices across agencies and countries for surveying multilingual populations. Patricia Goerman will present on the U.S. Census Bureau's history of testing non-English messages for use in written and verbal survey materials. She will discuss a doorstep messaging study currently underway and will discuss ways in which linguistically and culturally tailored messages might be coordinated across contacts and materials in a given survey operation. Charles Lau and colleagues will present on focus group research related to potential bias that can be introduced by third party presence during face-to-face interviews. This research focuses on work done in five African countries where third party presence is often a result of mistrust of researchers. Finally, Maichou Lor discusses the often-critical need for a third party support person among populations with lower literacy levels, such as elderly Hmong respondents in the U.S. She also discusses results of tailoring a survey instrument to include both multiple languages and visual cues for non-literate populations.

Maximizing Cultural Competence in Survey Data Collection

Alisú Schoua-Glusberg

Cultural competence in research is "the ability of researchers and research staff to provide high quality research that takes into account the culture and diversity of a population when developing research ideas, conducting research, and exploring applicability of research findings." (Harvard Catalyst 2009) In fieldwork, we need to keep cultural competence in mind at every step of the process, including preparing data collection tools, materials, and procedures. In this presentation we will discuss how to

bring cultural competence to these steps and what are some of the situations in which the quality of the data may be compromised by lack of understanding of the study participants' culture. For any specific population, culturally competent tools and materials for data collection must be prepared with an understanding of what information can be asked and answered in a survey in that culture. If the language of the group is not the same as the researchers', the tools and other respondent materials will need to be translated before pretesting them with members of the population. Approaching study participants or their households will require an understanding of the exposure and experience the population has with surveys, what will motivate them to participate or deter them from participation, within the framework of their cultural views. Both US and international fieldwork examples will be presented and discussed.

An Examination of Current Interviewer Training Practices for Multilingual Populations

Kathleen Kephart

As the United States becomes more globally connected, it has become crucial for major government agencies and survey practitioners to offer national survey materials in multiple languages. Information about immigrant populations is key for policy makers and data collection efforts must include these populations. Field interviewers working on U.S. Census Bureau surveys often encounter populations with Limited English Proficiency (LEP). In addition to not sharing a common language with the interviewer, LEP respondents may be less familiar with surveys such as the American Community Survey or the decennial census. There may also be cultural differences between the interviewer and household about greeting and linguistic conventions related to initiating conversations. These barriers may decrease the likelihood of household members responding, even when materials are available in a shared language. This challenge is not unique to the U.S. Census Bureau or other survey organizations in the United States. Increasingly diverse immigrant populations all over the world are leading to a need for interviewer training to handle this situation and optimize response. This paper will compare the current formal interviewer training practices of different institutions, such as non-profit organizations, international survey research organizations, academic institutions, and other governmental organizations to see what is common practice when encountering respondents who do not speak one of the primary languages of the country where they are being interviewed. I will perform a meta-analysis of current interviewer training practices available in the literature. This research will also attempt to examine the practices of government organizations outside of the U.S.

Development of Doorstep Introductory Survey Messages for Use across Languages: New Advancements towards Best Practices.

Patricia Goerman and Yazmín García Trejo

The U.S. Census Bureau has been engaged in research on messages to encourage survey participation of people who experience language barriers in the U.S. for much of the last decade. Research has included both written messages for materials such as letters and brochures and verbal messages for census interviewers and communications campaigns (Pan and Landreth, 2009; Boone, 2009; Bates and Pan, 2010; Pan and Lubkemann, 2010, Goerman et al. 2016). The types of messages tested include greetings, scripted and unscripted requests for survey participation, the mandatory nature of the Census and confidentiality of respondent data, among others. Over the years a number of qualitative and quantitative methods have been employed to study survey messages: field observation, expert review, focus groups and public opinion data (e.g. the Census Barriers, Attitudes and Motivators Surveys). In order to expand on the qualitative research related to messaging, researchers at the Census Bureau are currently conducting focus groups with speakers of Spanish, Chinese, Korean, Vietnamese, Russian and

Arabic to develop “best practices” in terms of the messages and other culturally appropriate practices most likely to encourage survey response in each language. This paper will provide a comprehensive historical overview of the findings related to various written and verbal messages that have been tested across languages for both written materials and verbal use. Our goal is to discuss methodological issues and preliminary findings from the current research on doorstep interviewer messages as well as next steps for future development of training and support materials for interviewers at the doorstep. This research will contribute to the discussion of ways in which messages can be coordinated across contacts and materials, both written and verbal, in order to increase survey response amongst hard-to-reach populations.

Bystander Presence during Face-to-Face Surveys: Perspectives from Interviewers from Four African Countries

Charles Q. Lau, Melissa Baker, and Clark Letterman

Face-to-face surveys are assumed to be conducted in private. In practice, however, other people (bystanders) are often present – particularly in low-income countries. Bystanders are an important feature of the cultural context that survey organizations must address to collect scientifically valid data. Bystanders can encourage respondents to edit their responses, potentially introducing measurement error. Existing research is quantitative in nature, and describes the number and types of bystanders present. But we have limited knowledge about interactions among the bystanders, respondents, and interviewers. For example, what motivates bystanders to observe the interview? What are the mechanisms through which bystanders influence respondents? How do interviewers attempt to minimize bystander influence? And how do bystander dynamics depend on interviewer and respondent’s characteristics such as gender? To investigate these questions, we conducted eight focus group discussions (FGDs) with interviewers and supervisors from Kenya, Nigeria, Tanzania, and Uganda. Participants discussed (1) what attracts bystanders to observe surveys; (2) how specifically bystanders lead respondents to change their answers; (3) how interviewers dissuade individuals from observing interviews or influencing responses. Our content analysis of transcribed FGDs paints a rich portrait of bystander interactions. We also consider how the cultural context of bystanders varies across countries. We find that negative views about survey teams encourage people to observe interviews, and make it difficult for interviewers to minimize influence of bystanders. Respondents generally view interviewers with skepticism, and sometimes mistakenly believe survey participation can produce benefits such as development aid to their community. Against this backdrop, interviewers face challenges in collecting accurate data. When bystanders are present, respondents tend to under-report socially undesirable behavior and over-report desirable behavior. Bystanders rarely help respondents answer factual questions. We provide guidance for interviewer training and survey introductions that acknowledges the cultural context of bystanders in each country, but also minimizes bys

ACASI-H: A Tool for Survey Completion in Non-literate and Non-English Speaking Older Hmong Adults Maichou Lor

As cultural and linguistic diversity continues to increase in the United States, researchers are challenged to expand their data collection techniques beyond a monocultural context. Health disparities are invisible, making it difficult to target interventions and to improve population health. There is an increasing number of older adult immigrants, who are non-literate in both their native language and English, live in a multigenerational household, and rely on their younger family helpers to complete their surveys when literacy is required. There are currently no known data collection strategies for use with this population. This presentation describes the development and feasibility testing of an adapted audio

computer assisted self-interviewing (ACASI) data collection tool, adding color labeled response categories and family helper assistance (ACASI-H). Sixty non-literate older Hmong (n=30) adults and their usual family helpers (n=30) comprised the sample. The instrument has simultaneous presentations of question items in two languages, both orally in Hmong and in English writing to accommodate the range of language proficiencies that were anticipated. Results revealed that older Hmong adults would not have been able to complete the survey without including a support person. Some family helpers were not very proficient in Hmong so having the simultaneous presentation in English was important. Color coded labeled response categories can have unanticipated consequences and should be used cautiously. An aural survey using color labeled response categories and including a family helper is promising and could be appropriate to engage respondents, who are non-literate in both their native language and English in survey interviews. This data collection tool is important and could be used for people who are not only from oral cultures, but also for people who are not highly literate in their native language. It could also facilitate collection of health data on populations we currently know very little about.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Panel Session on Panels (Panel) Organized by PANJAAPOR

PANJAAPOR's Panel Session on Panels

Yvonne E Shands SSRS
Michael Dennis

PANJAAPOR's Panel Session on Panels Panel Session Abstract for the 2017 AAPOR Conference, May 18-21, 2017 Proposed for the 2017 AAPOR Conference, May 18-21, 2017 We are all too aware of the current state of telephone interviewing—declining number of landlines, TCPA concerns, and increasing costs. In response, we have seen the emergence of both nonprobability and probability panels. In this era of rapid change, it is critical that researchers understand the core differences between these types of panels as well as the best fit for purpose for each. Considerations should include the need to balance practical concerns such as time and cost while carefully evaluating the methodological concerns of representativeness and measurement accuracy. This session focuses on current, actionable best practices in panel research. Experienced researchers from industry-leading organizations (e.g., Abt/SRBI, GfK, NORC) discuss how on-line probability panels fit into the overall research landscape. Topics include how probability-based panels originated, why they were created, and how researchers successfully employ them in their work. Panelists also provide insight into the unique aspects of probability panels including how they are maintained and the challenges associated with ensuring high data quality Co-proposers: PANJAAPOR, Abt/SRBI, GfK, NORC

Michael Dennis

Chintan Turakhia

Mansour Fahimi

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Surveying Physicians and Clinicians

The impact of incentive-type and mode-response options on cooperation: Evidence from experimental design in a physician survey

Gillian SteelFisher *Harvard School of Public Health*

Eran Ben-Porath *SSRS*

Linda Lomelino *SSRS*

Hannah Caporello *Harvard School of Public Health*

Two experiments were embedded in a survey of physicians who care for pregnant women and those planning pregnancies. The experiments involved a two-by-two factorial design: The first tested the impact of cash versus checks as a pre-incentive included with an invitation to participate. The second, tested the effect of offering multiple modes of response upon inviting respondents to participate (mail or online), compared with a single mode (mail only). Consistent with the literature on incentives for non-specialized respondents (i.e. households), the main hypotheses posited that: (1) cash incentives would yield higher cooperation than checks; and (2) offering multiple-response modes would yield lower cooperation than a single mode. The findings indicated support for both of these hypotheses. These were driven, in large part, by the markedly poorer cooperation rate where incentives were check-only and respondents could choose whether to complete the study online or by mail. The experimental conditions did not seem to correspond with differences in the substantive responses to the questionnaire. Beyond the particular findings, this paper considers the implication for best practices in recruiting physicians for high response-rate, opinion studies, along with the financial implications of design decisions such as incentive and mode of response.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Using Paradata to Improve Survey Administration

Estimation of Survey Cost Parameters Using Paradata

James Wagner *Univ. of Michigan*

In many survey situations, detailed cost parameters are difficult to estimate. This is especially true in survey involving interviewers. Overall costs may be easily estimated since interviewer hours, materials, and incentives are relatively easy to track. But costs at a more granular level – for example, hours spent travelling, identifying non-sample units, or engaged in other activities -- may be difficult to track. This occurs for a number of reasons. Often, cost information and paradata are collected in separate systems; or the cost information that is collected may not be at a sufficiently detailed level in order to evaluate the costs of particular subtasks. It might be possible to gather these cost data via a special study, but this is usually a very expensive approach. It may also be possible to ask for more detailed reporting from interviewers and other staff. However, this approach might lead to reduced efficiency. Instead, we propose to use regression models estimated from paradata in order to estimate detailed cost parameters related to interviewer effort. We will show examples of this method from the National Survey of Family Growth 2011-2018. This method was used to evaluate the costs of two treatment arms in an experimental study. The method is also used to monitor interviewer effort over the course of the field period.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

www.BetterWebSurveys.com: Increasing Data Quality for Web Surveys

Not This Again! Motivating Panelists to Achieve Daily Goals

Erin Wittkowski *Nielsen Company*

Arianne Buckley *Nielsen*

Participants in research requiring recurring or longitudinal measurement are asked to repeatedly engage in tasks that can be monotonous, repetitive, and even disruptive in nature. This can leave researchers facing inconsistent or declining cooperation rates over time. While incentives can improve panelist cooperation rates across waves, inducing panelist cooperation rates for recurring daily tasks can be particularly challenging. Nielsen PPM panelists currently receive incentives based on weekly compliance goals for cumulative time spent wearing a Portable People Meter. In July 2016, Nielsen conducted a study that explored whether daily compliance goals were as effective as weekly goals at incentivizing meter wear. For this test, panelists could earn an additional incentive on top of their standard monthly incentive if they met their specified compliance goals over a two week period. Panelists were randomly assigned to one of four groups: 1) a test group tasked with weekly compliance goals plus an opportunity for an additional multi-week bonus, 2) a test group tasked with daily compliance goals plus an additional multi-week bonus, 3) a test group tasked with daily compliance goals but no additional multi-week bonus, or 4) a control group that was not offered an additional incentive during the test period. The test results indicate that an incentive structure based around achieving consecutive daily goals can be equally, if not more, effective than a weekly-based incentive structure at improving panelist compliance. Additionally, the results showed that the addition of an extra incentive based on multi-week compliance also positively impacts panelist cooperation. The results of this test, along with the test design, will be detailed in this presentation.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Diversity: Attitudes and Measurements

Reconstructing Diversity: using polling archives to study diversity

Thomas Marshall *Univ. of Texas - Arlington*

Polling archives offer a potentially useful approach to understanding America's growing diversity. This presentation examines how well publicly-available polling archives can be used to study the opinions of special groups ("diversity groups") of public interest. Such groups are defined either by shared attitudes and beliefs (for example, birthers or Moslems), by demographics and background (dreamers; millennials; the very old); by behaviors or status (single-income, no kids (SINKS); LGBTs; the disabled; ex-convicts); or by some combination thereof (soccer moms, angry white men). Diversity groups vary widely in size and often number less than a tenth of all adults. Reconstructing diversity groups' opinions depends on whether and how often archived polls measure the group's characteristics, overall poll sample size, frequency of polls, the group's share of the population, specific related question wording, and how many attitudes are commonly measured across polls. This presentation addresses two questions. First, how often can diversity groups reliably be reconstructed from archived polls? Second, on what types of poll questions can diversity groups' opinions most frequently be measured? A dozen diversity groups are examined based on several publicly available poll archives.. For many diversity groups adequate poll samples are surprisingly difficult, and sometimes impossible to construct, even by aggregating from different polls. Specific examples, practical limitations, and useful research strategies are described and evaluated.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Got This on Your Calendar? Research on Events, Reference Periods and Dates

Time after time: Exploring the impact of fixed-start, variable-length reference periods

Maura Spiegelman *National Center for Education Statistics*

Reference periods are important both for maximizing respondent recall and balancing seasonal effects when collecting data on event counts. For some event types, respondents may naturally encode their experiences relative to a fixed starting point (e.g., the beginning of a school year), and survey questions that use this fixed starting point can yield more accurate responses than reference periods with variable starting points and fixed lengths (e.g., within the past two weeks). However, counts can only accumulate over time, so true values depend on when a question is answered; with fixed starting points, late responders are instructed to consider a longer reference period than early responders. This paper analyzes data from two surveys that use reference periods with fixed starting points and variable lengths: the 2010 School Survey on Crime and Safety (SSOCS), conducted February through June, and the 2012 National Household Education Surveys (NHES), conducted January through August. Both surveys ask respondents to count events that have occurred since a fixed starting point, for example, the number of unplanned fire alarms or days a child has been absent since the beginning of the school year. Responses will be examined both as straightforward counts and transformed to rates over a fixed, standardized reference period. Finally, similar early- and late-responders will be matched in order to determine whether response propensity explains differences in event counts or rates of events between these two groups. Results are expected to differ by survey item, for example, with conscientious early NHES responders indicating higher rates of attending school activities and lower rates of child absences than late responders, suggesting that transforming counts to rates may not be appropriate in all instances. This discussion of tradeoffs between recall, measurement error, and response propensity will be valuable to surveys that use fixed start, variable length reference periods.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Tracking the Election to Understand Trump's Win

RAND 2016 Presidential Election Panel Survey (PEPS) Polling Post-Mortem: What Went Wrong?

Michael Pollard *Rand Corporation - Santa Monica, CA*

Joshua Mendelsohn *RAND Corporation*

This presentation will use the longitudinal 2016 Presidential Election Panel Survey (PEPS) data in the RAND American Life Panel (ALP) to explore why national poll data missed the mark. The ALP consists of a panel of about 6,000 U.S. respondents ages 18 and older who regularly take surveys over the Internet. An advantage over most other Internet panels is that the respondents to the ALP need not have Internet access when they are initially recruited and thus the panel can be based on a probability sample of the U.S. population. The PEPS is uniquely positioned to conduct a polling post-mortem. It collected a comprehensive baseline survey in December 2015, including questions on party affiliation, primary voter choice, probabilistic polling, many topical issues (e.g., immigration, the economy, race/ethnic attitudes), head-to-head matches (e.g. Clinton vs. Trump), and candidate traits that matter. Respondents were re-interviewed throughout the election season for a total of six waves, including a post-election survey. Responses are also linked back to the respondents' original 2014 midterm and 2012 presidential election data where possible. Of significant importance, the data include longitudinal estimates of vote intention and participation likelihood, as well as information on what respondents ultimately did on Election Day, and why they did not vote, if they did not, which let us examine how the reported pre-vote intentions of respondents matched with what they actually did – and for whom the match was better or worse. In particular, we examine why Democratic voters turned up in fewer numbers than polls originally anticipated, and why Republican voters maintained recent levels of participation despite substantial disapproval of their candidate (Trump). All data collected are available for free to researchers. The ALP is also a service for researchers to field their own questionnaires and experiments.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Using Paradata to Improve Survey Administration

Exploring the Potential Use of Paradata Models to Inform Survey Extension Decisions

James Lawrence *US Census Bureau*

In recent years, surveys have struggled to achieve their past levels of overall response as one of the primary means of maintaining sample representativeness. It has therefore become increasingly important for survey managers carefully consider field interventions that are intended to protect the quality of the collected interview sample. For surveys with a short window of interviewing, this may require extending the deadline for completing interviews, so that additional responses can be collected. This paper presents the preliminary results of an ongoing project at the US Census Bureau to predict, during data collection, what a survey's response rate will be at its ordinary closeout time, so that a more informed, data-driven decision can be made regarding whether the survey needs to have its collection period extended. To create these models, we use readily available survey paradata, such as daily response rates, proportion of the collection sample still outstanding, and information about the type of contact attempts being made, so that we are able to incorporate the most up-to-date status of our field operations when the extension decision needs to be made. Because basic survey paradata is used in the creation of the models for this application, a similar approach may work well on other surveys. We will discuss a variety of models and the strengths and weaknesses of each, and we will discuss how useful survey managers can expect these sort of projections to be from a practical standpoint.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Diversity: Attitudes and Measurements

Leaving Room for Gender Diversity: Practical Lessons from a Blank Line in a Conservative State

Alian Kasabian *University of Nebraska-Lincoln's SBSRC*

Jenn Rutt *University of Nebraska-Lincoln's SBSRC*

Recent years have seen a rise in the visibility of gender minorities in the public eye, and an associated interest in data about this population. Although researchers have begun to study gender diversity, it has primarily been addressed through qualitative work, or smaller, regional studies. As of yet, benchmark surveys have not included measures of gender diversity in the general population. The September 2016 report from the Federal Interagency Working Group on Improving Measurement of Sexual Orientation and Gender Identity in Federal Surveys pointed out that there is a lack of consensus about how to measure gender diversity, and that more research is needed. We present the results of using a sex question that provides room for gender diversity by adding a blank space (which also allows for foolishness) while still meeting the needs of local research clients. It was favorably received by both gender minority and non-minority respondents in testing, with evidence of improved engagement with surveys that provide room for minority identification. We draw on data collected in a conservative, Midwest state from over 3,000 paper surveys implemented across different populations, 39 cognitive interviews, and direct questions about the format which is asked of respondents via MTurk to illustrate the impact of using a question like this. Although not ideal in terms of gender measurement, the proposed measure fits efficiently into diverse types of surveys, provides gender measurement room for those that want it, and the majority of respondents used it “correctly” – that is, without foolishness. The paper discusses the practical lessons – both positive and negative – learned from our data.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Using Paradata to Improve Survey Administration

Using Timing Metadata for Data Validation and Quality Control: The NLSY97 Case Study

Ilana Ventura *NORC at the Univ of Chicago - Chicago, IL*

A. Rupa Datta *NORC at the University of Chicago*

The National Longitudinal Survey of Youth 1997 (NLSY97), is a longitudinal nationally representative study that gathers information on the labor market experiences of American men and women. The initial cohort included 9,000 youths who were between the ages of 12-18 in 1997. The NSLY is sponsored and funded by the U.S. Bureau of Labor Statistics (BLS) of the U.S. Department of Labor, with data collection by NORC at the University of Chicago. NORC has developed several methods for data quality review and data validation for the NLSY. In the 2015 survey round for the NLSY97, which used CATI and CAPI for data collection, time spent by respondents on each question was recorded. This work focuses on how the NLSY97 used timing metadata for data validation through comparisons between cases, interviewers and modes, as well as to compare case timings with prior interviews. Analysis of item-level timing can help ensure that definitions and consent language are read verbatim, and can also help identify questions that may not be performing consistently across respondents. Information on the timing of transitions, such as during transitions between CAPI and SAQ portions of the survey, or when starting a locator section, can assist in identifying issues with interviewers or with the protocol. As we prepare for mode transition, timing data at the section and subsection level can help us understand differences between the modes with regard to places for questionnaire improvement. Finally, overall questionnaire timing, if too long, can indicate a poorly functioning interview or, if too short, an interviewer issue. This analysis methodology allows for close monitoring of the data collection process, and adds a layer of automated quality control. This work will inform survey researchers looking to improve quality control measures by taking advantage of timing metadata.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Diversity: Attitudes and Measurements

Attitudes Towards Workplace Diversity Policies: Race, Gender, and Discrimination as a Source of Inequality

William J. Scarborough *University of Illinois at Chicago*

William J. Scarborough *University of Illinois at Chicago*

Danny Lambouthis *University of Illinois at Chicago*

Allyson Holbrook *University of Illinois - Chicago*

Since the 1980s, workplaces across the U.S. have increasingly adopted diversity policies aimed at improving the representation of women and racial minorities (Dobbin 2009; Kalev, Kelly, and Dobbin 2006). The structure of these policies varies across firms and include such measures as mandatory diversity training, mentorship programs, and diversity offices. Despite the widespread presence of workplace diversity policies, we know little about peoples' attitudes towards them. Most previous research has focused, instead, on attitudes toward affirmative action. However, modern workplaces rarely use the language or model of affirmative action when instituting diversity policies (Bielby 2000). In this study, we use a unique survey instrument to investigate peoples' attitudes towards eight different workplace diversity policies. We first examine whether peoples' support for workplace policies vary depending on the policy's characteristics. Next, we look for differences between race and gender groups in support for diversity policies. We also examine whether support differs depending on whether the policies are targeted towards increasing the representation of women or racial minorities. Finally, we examine whether beliefs about the sources of race/gender inequality mediate the effects of race, gender, and policy. Our preliminary findings indicate that individuals' attitudes toward policies do not differ depending on the characteristics of the policy. Women, African Americans, and Latina/os are more supportive of diversity policies than whites and men, and between 30 and 40 percent of these race/gender differences can be explained by differences in the extent respondents believe discrimination causes gender/race inequality. Furthermore, whites show more support for diversity policies when they are aimed at increasing the representation of women as opposed to racial minorities. About 40 percent of this difference is because whites feel that gender inequality results from discrimination to a greater extent than they feel that racial inequality results from discrimination.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Tracking the Election to Understand Trump's Win

A developmental analysis of Trump voters

Jon Miller *University of Michigan*

The Longitudinal Study of American Life (LSAL) has been following a national probability sample of 4,000+ students for the last 29 years, beginning when the students were in grade 7 or grade 10 in a public school in the United States. Over the last 29 years, approximately 12,000 variables have been collected from each participant reflecting a wide array of educational, occupational, political, and personal attributes. These young adults are now 41 to 44 years of age and live in all 50 states. The study was funded by the National Science Foundation for 25 years and it now supported by an R-01 grant from the National Institute on Aging. This paper will utilize this extraordinary longitudinal record to characterize early and mid-life political involvement, partisanship, policy attitudes, family and economic factors, and related variables on the vote decision in 2016. All of the reported voting activity will be verified through Catalist. In addition to responses from participants, all of the participant addresses in the study have been geo-coded and GIS software will be utilized to characterize the neighbors and communities in which participants currently live. The data collection is underway at the time of this proposal submission (less than a week after the election). It is obviously impossible to describe the results of this year's annual survey, but it should provide interesting new insights into the factors that produced votes for Donald Trump.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Tracking the Election to Understand Trump's Win

**What We Learned from Conducting the Experimental USC Dornsife / Los Angeles Times 2016 Election
“Daybreak” Poll**

Jill E. Darling *University of Southern California (USC) Center for Economic and Social Research*

Arie Kapteyn *University of Southern California*

Erik Meijer *University of Southern California*

Tania Gutsche *University of Southern California*

In 2016, the USC Center for Economic and Social Research Understanding's America Study (UAS), conducted an experimental poll using its probability-based internet panel. We used a probabilistic approach pioneered at RAND by Delavande & Manski (2010), and successfully used to forecast the 2012 presidential election (Gutsche, Kapteyn, Meijer, & Weerman, 2014; Kapteyn, Meijer, & Weerman, 2012). From July 4 to November 7, respondents in the UAS election panel answered vote questions on an assigned day once per week. We calculated the ratio of vote percentage to turnout percentage for each candidate and estimates were presented in the form of online charts updated nightly as 7-day rolling averages. The Daybreak poll's estimate often showed Donald Trump ahead when others had a Clinton lead, which engendered heated responses. As this work is part of an ongoing exploration of the utility of methods that may help address problems facing the field of election polling, we made our data and methods available to other researchers for analysis during the election season and after. We will give a brief overview of the poll's alternative methodology; provide some new information from a post-election analysis of the predictive power of our probabilistic questions, and share what may be surprising findings from our re-examination of our weighting model. We will also touch on our personal experience with conducting an outlier experimental poll in a divisive election year and present ideas for how our methods may contribute to the field of election polling going forward

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Got This on Your Calendar? Research on Events, Reference Periods and Dates
Non-Specific Reference Periods in Survey Questions: Understanding Survey Responses about Behaviors in a "Typical Week" Versus "Last Week"

Matthew Virgile *U.S. Census Bureau - Washington, DC*

Jonathan Katz *U.S. Census Bureau*

Jasmine Luck *U.S. Census Bureau*

Across surveys, behavioral questions vary in specificity of reference periods. Some surveys ask for information within a specific period, such as “last week,” and/or providing exact dates (e.g. “Since October 1, 2016...”). Other surveys, however, ask for information within a non-specific period, such as “in a typical week.” First, we compare items on national surveys that use specific versus non-specific reference periods. Then we present findings from exploratory research, including an Internet survey with questions about the frequency of various behaviors. Separate questions include non-specific reference periods such as “in a typical week” or specific reference periods such as “last week.” These questions are followed by probes about respondents’ understanding of the reference periods in regards to each behavior. We discuss how respondents formulate their answers to these items, including when they do not have a “typical” week in regards to one or more behaviors.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Surveying Physicians and Clinicians

Can we increase the impact of prepaid incentives using loss aversion theory?

Nikkilyn Morrison *Mathematica Policy Research*

Jared Coopersmith *Mathematica Policy Research*

Nancy Duda *Mathematica Policy Research*

Studies have shown that prepaid incentives are effective in raising physicians' and other health care professionals' responses to mail surveys. This is explained by the norm of reciprocity – receipt of a benefit invokes doing something in return. However, loss aversion theory suggests people may pay more attention to a loss than a benefit. When offered the opportunity to avoid a loss or achieve a gain of equivalent value, people will focus on the loss over the gain, and act to avoid the loss. In a mail-only survey of primary care clinicians and practice staff, we implemented an experiment to test whether deploying loss aversion theory in follow-up communications about a prepaid incentive can boost response rates. We hypothesize communications that emphasize a potential loss will draw more attention to the survey than those just highlighting the equivalent gain. For the evaluation of the Comprehensive Primary Care Initiative, sponsored by the Centers for Medicare and Medicaid Services, we conducted this experiment with 1,800 primary care clinicians and 2,600 staff. Clinicians received a \$100 check and staff received a \$25 check in the first questionnaire mailing. For the treatment group in this experiment, the three follow-up questionnaire mailings over four months included a reminder about the prepaid incentive check, language emphasizing the potential loss of the check, and an easy method to replace the check. The treatment questionnaire included a checkbox on the cover which respondents could mark to request a replacement check. The control group follow-up mailings simply reminded nonrespondents that they had already received the incentive, and their questionnaire did not have a checkbox on the cover. We examine whether the new text and checkbox on the questionnaire yield higher response rates, affect cashing of the prepaid checks, and increase the cost of re-issuing the replacement checks.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Surveying Physicians and Clinicians

Examining the Effect of Declining Response Rates on Nonresponse Bias

Emily Geisen *RTI International*

Rebecca Powell *RTI International*

Marshica Stanley Kurtz *RTI International*

Murrey Olmsted *RTI International*

Joe Murphy *RTI International*

Nonresponse bias is determined by both the nonresponse rate and the difference between nonrespondents and respondents to the survey (Groves, 1989). Declining response rates have a negative effect on precision and on accuracy if differences between nonrespondents and respondents are magnified (Groves, 2008). In this paper, we examine the relationship between response rates and nonresponse bias on a survey over time. Since 2005, RTI has conducted a national, cross-sectional survey of physicians each year as part of U.S. News & World Report's Best Hospitals rankings. Over this time, response rates declined by 21 points from 47% in 2005 to 26% in 2016. The methodology varied slightly by survey year, but mail was the predominant mode in all years. In 2005, we found that there were modest amounts of bias for gender (women were less likely to respond) but not region or urbanicity. Although there were differences in response rates by specialty, lower response rates were not associated with greater bias for gender, region or urbanicity (McFarlane et al., 2007). The current paper extends this analysis to data from the past ten years of surveys, to determine the impact that declining response rates have had on nonresponse bias. We analyze whether demographic factors available on the frame (gender, region, urbanicity, age, and hospital affiliation) affect the likelihood and timing of survey response. Respondent outcome measures are then analyzed with respect to the demographic variables. The outcome measures of interest are (1) nominating a hospital they are affiliated with, (2) nominating a top hospital in their specialty, and (3) nominating only hospitals in their region. This paper will describe the factors associated with nonresponse each year, the direction of nonresponse bias, and the relationship between response rates and nonresponse bias over time.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

More Effects a la "Mode"

Examining Mode Effects in a National Survey of Teachers and Principals

Alina N Kline *US Census Bureau*

Joanna FaneLineback *US Census Bureau*

In this presentation, we discuss the effect of response mode (paper vs. internet) on data quality in the National Teacher and Principal Survey (NTPS). Many surveys that produce official statistics have moved or are investigating the move to multi-mode data collection methods that include an internet component. The NTPS is one such survey. Actually a series of three surveys, one for each schools, school principals and teachers, the NTPS provides detailed statistics on U.S. elementary and secondary public schools, principals and teachers. Prior to the 2015-2016 NTPS, the information collected by the NTPS was collected as part of the Schools and Staffing Survey, which primarily was a self-administered, paper survey. The 2015-2016 NTPS teacher survey was a multi-mode survey, offering both paper and internet response options, while the school principal survey offered paper response except to 1,000 randomly selected principals who were offered internet response. Using survey response as a measure of data quality, we compared the differences between modes. For the teacher survey comparisons, we used propensity score weighting adjustments. This quasi-experimental approach was possible because we had access to a rich set of frame data. For the school principal survey comparisons, we used regression and time-to-event models. These findings will be used in conjunction with internet paradata and cost findings to inform data collection strategies for future survey cycles.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Surveying Physicians and Clinicians

MEPS Medical Provider Component Medical Organizations Survey: Is a Linked Survey Strategy More Successful In Getting Office Based Medical Providers to Participate in a Survey?

Marie N. Stagnitti *Agency for Healthcare Research and Quality*
Kathryn Dowd *RTI International*

The Medical Expenditure Panel Survey Household Component (MEPS-HC) is an ongoing household survey that yields national estimates of various health care metrics including health care use, expenditures, and insurance coverage. The MEPS Medical Provider Component (MEPS-MPC) collects information from medical providers providing care to the MEPS households. The provider data are an invaluable complement to the household reported data. Often more detailed and accurate, the provider data serve as the gold standard for MEPS expenditure estimates and are the source for MEPS expenditure imputations. Because of increased demand for data on organizational characteristics of providers and/or health care practices, the Robert Wood Johnson Foundation has sponsored a Medical Organization Survey (MEPS-MOS) which collects this type of data from a subset of MEPS-MPC providers. Physician surveys are known to be difficult to execute and get low response rates. The linked MEPS-MOS survey approach examines if this unique data collection strategy is easier to execute and results in higher response rates. Results of this survey effort will provide critical information to future efforts in provider and medical practice data collection. This paper will present underlying design considerations of the MOS instrument development and data collection strategy. It will include discussion on successes and challenges of the linked survey approach and will present response rates by question and respondent category (i.e., the respondent's role in the provider organization). Item non-response and analytic potential of the data will be discussed. Preliminary findings show an overall response rate of 79 percent at the person-provider and practice levels with phone data collection being the most prevalent mode (90 percent) compared to web (4 percent), and fax/mail (6 percent) modes.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Got This on Your Calendar? Research on Events, Reference Periods and Dates
Improving Reporting Date of Victimization Events in Mail Surveys

Douglas Williams *Westat*

J. Michael Brick *Westat*

W. Sherman Edwards *Westat*

Pamela Giambo *Westat*

Lynn Langton *Bureau of Justice Statistics*

Survey respondents find it difficult recalling biographical events when events are frequent or temporally distant. With frequent non-routine events it is difficult to distinguish the details for each event such as the date. For temporally distant events it is well-known that respondents have difficulty determining if an event occurred before or after an arbitrary reference period. These issues are important in the National Crime Victimization Survey, a face-to-face survey of the Bureau of Justice Statistics. It collects details about criminal victimization including the date the victimization occurred. Only those victimizations with dates in the reference period are eligible. These issues are even more problematic in a lower cost mail survey called the National Crime Victimization Survey Companion Study (NCVS-CS). This survey was field tested in 2015, and then modified and fielded again in 2016. In 2015, we observed increasing levels of item nonresponse for victimization dates that are later in time. Questionnaire design theory recommends placing date items at the end of the series of items with the hope that the intervening questions improve recall. The field test also fielded a questionnaire version that placed these items at the end and found even higher item nonresponse. This paper examines the effect of questionnaire changes made to the 2016 instrument to reduce item nonresponse, with emphasis on a categorical question added to collect victimization date. This enables us to examine the hypothesis that respondents know whether the victimization occurred within the reference period, but are unwilling estimate a specific date. We compare the accuracy of the categorical item by comparing it to the month and year reported. We also examine how the placement of this question impacts victimization reporting and item nonresponse. Finally, we look at item nonresponse by victimization severity as more severe victimizations are likely to be more salient.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

More Effects a la "Mode"

Surveying adolescent mothers: Examining self-reports to sensitive questions in different modes

Jennifer Walzer *Mathematica Policy Research*

Emily Weaver *Mathematica Policy Research*

Researchers often use mixed-mode data collection designs to reduce non-response bias, control costs, and increase response rates (Dillman et al, 2014). This can be particularly important when studying hard to reach respondents, such as low-income and mobile populations. However, while mixing the modes of data collection during a single wave may improve external validity, it also increases the potential for measurement error due to mode effects. The error introduced by these mode effects is most pronounced on sensitive questions (Gnambs & Kaspar, 2015). For the first follow up survey, 12 months post baseline, in an evaluation of programs for adolescent mothers (n=1500) we use a mixed-mode data collection design. Respondents are first offered a web mode, followed by CATI, and lastly an in-person field locator provides a device for the respondent to complete by web. The survey collects data on educational attainment, contraceptive knowledge, activities, including sexual activity, drug and alcohol use, and pregnancy history. This paper will compare differences by mode in self-reported behaviors on sensitive questions in an adolescent population, controlling for age and other demographics as needed. We will explore differences by the topic area of the questions and if differences are noted, we will explore if they are mitigated by the age of the adolescent. Research suggests that compared to older adolescents, younger adolescents may be more prone to social desirability bias (De Leeuw et al. 2004). This could increase measurement error on sensitive questions across completion mode for younger respondents. We are interested in better understanding this subpopulation of adolescent mothers, and the extent to which efforts to gain their participation through multi-modes affect key measures. Results will be generalized to provide recommendations on how to reduce mode effects and social desirability bias on sensitive questions in multi-mode surveys.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Using Paradata to Improve Survey Administration

Using Paradata to Measure Respondent Engagement

Margaret L. Hudson *University of Michigan*

Andrew L. Hupp *University of Michigan*

Heather M. Schroeder *University of Michigan*

Andrew D. Piskorowski *University of Michigan*

Understanding how participants engage with requests to participate in a web survey can provide survey practitioners with useful information on how to optimally reach their intended targets. This presentation uses data from two longitudinal studies from different populations. Both studies utilize a web beacon to gather paradata on participants engagement with email communications. Data generated include the dates and times emails are sent, the date(s), time(s), and platforms they are viewed. This information is combined with other paradata including the platform used to access and potentially complete the survey. The date(s) and time(s) these events occur allow a view into participant engagement. These data may allow survey practitioners to focus on engagement patterns and identify optimal times to email participants to maximize response rates. In addition to identifying potential patterns of engagement, the beacon provides information that can be used to identify obsolete email addresses as they become out-of-date over time. Many participants have provided multiple addresses from different email domains. While there may be a nominal monetary cost to sending emails to the same person with multiple addresses, there may be adverse side-effects such as raising spam scores or aggravating participants. Beacon information could be utilized to identify a primary email address rather than sending to all email addresses. This paper will discuss email opening rates, lag times from when emails are sent to when they are viewed, platforms used to view emails, and limitations of the beacon.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Using Paradata to Improve Survey Administration

Using Audit Trails to Support Questionnaire Design Improvements

Renee M Gindi *National Center for Health Statistics*

Carla Zelaya *National Center for Health Statistics*

Keystroke-level information collected in survey audit trails can be used to help questionnaire designers improve survey instruments. In this presentation, we share two examples of using audit trail files to improve data quality in the National Health Interview Survey (NHIS). In the first example, we use audit trail files to examine how interviewers were using a new survey feature that asked them to fix discrepancies in reported family income and summed personal earnings. Rather than using the feature as intended, many interviewers purposely overwrote original data with new values making discrepancies difficult to analyze. In order to determine the extent to which the feature may have changed income reporting in the 2011 and 2012 NHIS, we used audit trail data to reconstruct original values. We show how automated code can be used to search for discrepant cases and evaluate the utility of the new survey feature. In the second example, we use audit trail files to evaluate the utility of introductory survey questions aimed at improving the accuracy of the household roster. Current instrument programming does not allow for a simple evaluation of these items, because the original responses are overwritten by edited data. In order to determine the utility of each of these items in correctly identifying usual household residents in the 2015 NHIS, we used audit trail data to reconstruct the original responses. For both examples, we show how automated programs can assist in determining the value of questions and survey features. We discuss the advantages of using audit trail data to understand the use of different survey features, and we provide practical guidance on the use of audit trails.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Surveying Physicians and Clinicians

Maximizing Response Rates and Reducing Costs for Physician Surveys: When Should Pre-Incentive Checks be Sent?

Robyn Rapoport *SSRS*

Michelle Doty *The Commonwealth Fund*

Declining response rates in physician surveys have prompted researchers to conduct experiments aimed at reducing costs and improving response among these sought-after professionals. Numerous studies have demonstrated that monetary pre-incentives help induce cooperation and have been associated with meaningfully higher response rates than promised incentives (e.g., Delnevo et al 2004). In addition, researchers have investigated different design strategies (e.g., VanGeest et al 2007) and have noted the benefits of including paper surveys and mixed-mode options to improve response rates among physicians. This paper contributes to research on maximizing responses rates and reducing costs in physician surveys and reports on findings from an experiment included in a multi-mode survey fielded in 2015 with primary care providers (i.e., GPs, Internists, and Pediatricians) to determine whether providing a \$25 check in the pre-notification mailing (with the web link) is more likely to result in a completed interview compared with providing the check in the second mailing, along with both the web link and a paper survey. We will also investigate which design is most beneficial in terms of the cost per interview. Finally, we report on the prevalence of ‘cheaters’ – those who cashed the pre-incentive check but did not complete the survey vs. ‘Good Samaritans’ – those who completed the survey but chose not to cash their prepaid incentive check among physicians exposed to the different conditions. Findings to be discussed include the following: Response rate is higher among those that got their incentive with the paper survey (second mailing). A larger portion of those in the experimental condition completed the survey online compared with those that received their incentive in the second mailing (55% vs. 24%). PCPs in the experimental condition are more likely to be both cheaters and Good Samaritans...but the Good Samaritans more than counterbalance the cheaters.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Diversity: Attitudes and Measurements

Social Construction and White Attitudes Toward Multicultural Issues and Support for Presidential Candidates

Robert W. Oldendick *University of South Carolina*

Monique L. Lyle *University of South Carolina*

As the United States becomes increasingly racially and ethnically diverse, the role that racial attitudes play in support for public policies as well as in voting behavior becomes increasingly important. Using a social construction differentials approach, this study employs data from the 1992 through 2012 American National Election Studies to investigate cross-time changes in the cognitive images that whites hold of racial/ethnic minority groups in comparison to their own race. Previous research on this topic has shown that these cognitive images have a significant effect on attitudes toward group-based issues. This investigation extends this research through a period of heightened racial awareness in the United States, during which this country elected its first black president, to examine changes in whites' social construction of other groups during this time and how these changes may have varied across white subgroups, such as by education level or age. In addition, this research expands the scope of this investigation beyond group-based issues to include the impact of social construction on evaluations of ostensibly non-racial policy issues and candidates for the Presidency. Cross-time changes in whites' social construction of various groups may have contributed to the increasing polarization in public policy attitudes on a number of issues, including preferences for Presidential candidates.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

More Effects a la "Mode"

Disentangling Measurement Effects from Selection Effects in a Sequential Mixed-Mode Experiment

Rebecca Medway *American Institutes for Research*

Mengmeng Zhang *American Institutes for Research*

Cameron McPhee *American Institutes for Research*

Michael Jackson *American Institutes for Research - Washington, DC*

Mixed-mode data collection is an attractive option due to its potential ability to increase response rates, reduce nonresponse error, and reduce survey costs as compared to single-mode surveys; however, data collected in diverse modes also tend to have different measurement error properties. In particular, we may observe mode-driven “measurement effects” – that is, a difference in measurement error across modes caused by respondents answering the same question differently depending on the mode in which they respond. These measurement effects are potentially problematic because they suggest that respondents in each mode may be interpreting the survey questions differently, challenging the ability to confidently combine responses from these modes into a single data set. However, measurement effects can be difficult to estimate since they tend to be confounded with selection effects - differences between the characteristics of the respondents to the two modes. In this presentation, we will utilize two analytic approaches ((1) mixed-mode calibration and (2) extended mixed-mode comparison) to disentangle measurement effects from selection effects in the 2016 National Household Education Survey (NHES) data. This administration included an experiment, in which a random sample of the sampled households was assigned to a sequential mixed-mode condition that gave them the opportunity to complete the survey either on the web or, later, on paper (35,000 households). The remaining cases were assigned to a single-mode, mail-only condition, in which they received the standard NHES mail survey protocol (171,000 households). Disentangling selection and measurement effects allows us to assess both the extent to which the mixed-mode condition brought in respondents who differed on key survey topics from those who responded in a mail-only administration (selection effects), and the extent to which the measurement of these topics varied by mode of response (measurement effects).

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

www.BetterWebSurveys.com: Increasing Data Quality for Web Surveys

Gamifying incentives: A case study

Ting Yan *Westat*
Janice Machado *Westat*
Andrew Heller *Westat*
Aaron Maitland *NCHS*
John Kirlin *USDA*
Erika Bonilla *Westat*

Gamification is shown to improve user experience, boost user encouragement, and increase user loyalty (Richter, Raban, and Rafaeli, 2015). It has been increasingly applied to marketing (Richter, Raban, and Rafaeli, 2015), education (Richter, Raban, and Rafaeli, 2015), and even survey research (Puleston, 2011). One application of gamification in surveys is gamifying incentives. Only one pilot study empirically examined the effects of gamifying incentives and found that the gaming mechanics have a significantly positive impact on respondent behaviors. This paper discusses an incentive scheme guided by the gamification theory. Instead of providing the same amount of incentives to everybody, respondents are incentivized based on their behavior; they cumulate incentives only by fulfilling a required survey task. Following the gamification theory, we send daily notifications to respondents via text or e-mail (depending on respondent preference) of the amount of incentives they have accumulated. In addition, every time respondents complete a task, the amount of incentives accumulated is displayed to the respondent. This paper examines the effectiveness of this incentive gamification scheme on respondent participation and behaviors.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

More Effects a la "Mode"

How Much Does Mixed Mode Increase Response Rates? Evidence from a Randomized Mode Experiment in Hospitals

Marc Elliott *RAND Corporation*

Katrin Hambarsoomian *RAND Corporation*

William Lehrman *Centers for Medicare & Medicaid Services*

Laura Giordano *Health Services Advisory Group*

Megan Beckett *RAND Corporation*

Julie Brown *RAND Corporation*

Elizabeth Goldstein *Centers for Medicare and Medicaid Services*

Sequential mixed mode (initiating data collection in one mode, then pursuing non-respondents by a second mode) is now widely recognized for improving response rates and representativeness compared to single-mode approaches. However, some organizations are reluctant to incur additional cost and complexity without knowing how much response rates (RRs) would increase. We used data from a 2016 Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey mode experiment to (1) estimate the effect on RR of switching from more commonly used single mode protocols (Mail Only and Telephone Only) to a sequential Mixed Mode protocol, and (2) to identify which hospitals and age groups are most likely to benefit. 17,415 patients from 50 hospitals were randomized in equal numbers to (1) Mail Only, (2) Telephone Only; or (3) Mixed Mode (mail then telephone follow-up). RRs were predicted from hospital random effects, survey mode, and patient age in mixed-effect logistic regression models. For all age groups, RRs were highest for Mixed Mode and lowest for Mail Only mode. Mail Only RRs were lowest for ages 18-24 (7%) and ~30% for ages 65+. Telephone Only RRs were ~24% for ages 18-24, increasing to ~35% by ages 55-84. Mixed Mode RRs were ~30% at age 18-24, increasing to ~60% by ages 65-84. Poor hospital-level RRs by Mail Only strongly predicted greater gains from Mixed Mode. For example, a hospital with a 15% Mail Only RR has a predicted Mixed Mode RR >40% (with >25% in telephone follow-up). In such cases, Mixed Mode may be less expensive than the much larger Mail Only sample needed to achieve HCAHPS' required 300 annual completes. It may be that hospitals with low Mail Only RRs especially benefit from Mixed Mode because their patients may be disproportionately young, highly mobile, or with low literacy.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

More Effects a la "Mode"

Emergency Department Patient Experience of Care Survey in the Discharged to Community Setting – A Randomized Feasibility Study

Layla Parast *RAND Corp*

Layla Parast *RAND Corp.*

Megan Mathews *RAND Corp.*

Anagha Tolpadi *RAND Corp.*

Marc Elliott *RAND Corp.*

Elizabeth Flow-Delwiche *Centers for Medicare and Medicaid Services*

Kirsten Becker *RAND Corp.*

The Emergency Department Patient Experience of Care (EDPEC) Discharged to Community (DTC) Survey, developed by the Centers for Medicare and Medicaid Services, was designed to measure the experiences of patients with emergency room care. Results from a pilot test of this survey demonstrated low response rates when a mail only mode was used for survey administration. Here, we describe the results of the 2016 EDPEC DTC Feasibility Test which aimed to test the feasibility of different novel modes of survey administration in this unique setting. In this feasibility test, sampled discharges were randomly assigned to one of five modes: on-site distribution in the emergency department (ED), mail notification of web survey, email notification of web survey, enhanced mixed mode (email notification with mail follow-up and then with telephone follow-up), and standard mixed mode (mail with telephone follow-up). Feasibility was assessed in terms of response rates by mode, level of required involvement by hospital staff, and email capture rates. Approximately 4,017 patients discharged from eight EDs across the United States were sampled and randomized. There were significant differences in response rates between survey modes, with enhanced mixed mode being the highest (30.42%, AAPOR RR1), and mail notification of web mode being the lowest (0.76%). Email capture rates ranged from 3.1% to 65% by hospital; the odds of having an available email address were significantly lower for older patients and for male patients. Email notification of web mode respondents were significantly more likely to report poorer experience compared to standard mixed mode respondents; enhanced mixed mode respondents were significantly more likely to report better experiences than standard mixed mode respondents. Results from this feasibility test showed that on-site distribution would not be feasible, while enhanced mixed mode was identified as a promising novel mode for potential use in national implementation.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

www.BetterWebSurveys.com: Increasing Data Quality for Web Surveys
Using Survey Responses to Improve Retention in a Mobile Panel

Jana Dodson *Nielsen*
Ally Glerum *Nielsen*
Lanie Anton *Nielsen*

High panelist attrition rates are a persistent concern among survey researchers, particularly in longitudinal panel studies (Ribisl, et al., 1996). Understanding attrition and ultimately implementing strategies to reduce it will lead to survey cost reduction, data quality improvement, and enhanced panelist experience. While there is substantial research regarding strategies for reducing attrition in panel surveys, little has been done to tailor these strategies towards mobile panelists, despite having significantly higher attrition rates than other panelists (Fuchs, 2012). Using data from a continuous mobile Nielsen panel of approximately 40,000 participants nationwide, we examine the link between survey data quality and retention, as well as propose methods for promoting retention among panelists providing higher quality data. In particular, we target two primary survey response patterns that tend to produce lower quality data: a) “speeding,” or completing the survey significantly faster than expected, and b) “straight-lining,” or responding to a series of questions in the same place on a rating scale. Both behavior patterns indicate that the survey questions were answered insincerely or carelessly, threatening the integrity of the survey. The identification of such behaviors in initial surveys may allow us to predict attrition and replace panelists likely to produce lower quality data early on, thereby improving the cost effectiveness of administering the survey. We also look for an inflection point during panel membership after which attrition rates tend to drop precipitously, and panelists become significantly more likely to remain long-term. The study of these behavior patterns allows us to explore new methods for optimizing the distribution of incentives. Results from these analyses provide a new perspective to an extensive body of research regarding panelist attrition by focusing specifically on mobile panelist survey response behavior, gearing toward maintaining a stable and cost-effective panel with high quality data.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

www.BetterWebSurveys.com: Increasing Data Quality for Web Surveys
Leveraging Survey Experiments to Assess Respondent Commitments to Quality
David Vannette *Stanford University*

Concerns that survey respondents may not always provide thoughtful answers to questions have plagued researchers for decades. A large body of research has examined when and how respondents will provide low-quality responses. In fact, the theory of survey satisficing, which describes a set of respondent strategies for avoiding providing effortful responses, is one of the most prominent theories in survey methodology. A key factor that predicts when respondents will provide low-quality data is low respondent motivation. However, research examining strategies for increasing respondent motivation to provide high-quality data is sparse and much of the research that does exist focuses on the ancillary effects of monetary incentives to participate on data quality. This paper seeks to add to the literature on respondent motivation and data quality by assessing the effectiveness of soliciting a respondent commitment to provide high quality data via thoughtful answers, a method first proposed in the 1970s by Charles Cannell. The study design leverages a series of large-scale web survey experiments performed on diverse samples of respondents from around the world. The results indicate that respondent commitments do affect respondent behavior during survey completion and may influence data quality. The findings of this research will be of value to all survey researchers that are concerned with response quality.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Got This on Your Calendar? Research on Events, Reference Periods and Dates
Timing of Transitions in the Event History Calendar of the 2014 SIPP Panel

Shelley Irving *U.S. Census Bureau*
Matthew Marlay *U.S. Census Bureau*

In previous SIPP Panels, respondents were interviewed every four months. Under the newly redesigned 2014 SIPP Panel, respondents are visited early in the interview year and asked questions about the preceding calendar year (the reference year). The Event History Calendar (EHC) was developed to facilitate respondent recall over the longer reference period and allows for integrated reporting across the domains of residency, marital history, educational enrollment, employment/time not working, program participation (SSI, SNAP, TANF, WIC, and GA), and health insurance coverage. Data in the EHC are entered as spells and detailed information about each spell is then collected. For example, a respondent reported that he was employed at his current job from June 2013 to the present, and provides detailed information about this job. Next, the respondent reported that he worked at a different job from January 2013 to June 2013 and then provides detailed information about this other job. This paper is an evaluation of the timing of transitions within the EHC. Within topics, we are interested in the number and length of reported spells. For instance, do respondents tend to report a single spell spanning the entire reference year? This might make sense for a topic such as SSI receipt, but make less sense for school enrollment, where we usually expect to see a break during the summer months. Across topics, we want to observe whether there are interdependencies in topics. For example, is becoming employed correlated with a change in health insurance coverage status? Are transitions from employment to non-employment accompanied by changes in program participation status? Do people tend to move around the time that we see changes in educational enrollment, employment, or marital status?

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Tracking the Election to Understand Trump's Win

A 2016 Election Polling Post-mortem: The ABC News/Washington Post Tracking Poll

Gregory Holyk *Langer Research Associates*

Gary Langer *Langer Research Associates*

Chad Kiewiet De Jonge *Langer Research Associates*

Scott Clement *The Washington Post*

The ABC News/Washington Post tracking poll, conducted Oct. 20-Nov. 6, 2016, via national random-digit-dialed telephone interviews, provided an extensive examination of likely voters' views on candidate attributes and issues, divisions among groups and vote preferences. The poll's final estimate, in which our reports characterized the race having no clear leader, understated Hillary Clinton's support by 1 percentage point and Donald Trump's by 4 points, as compared with currently available vote totals. We are extensively reviewing and will report on apparent contributors to the imprecision in the Trump estimate. Our evaluation will include comparisons to national exit poll data and Current Population Survey voter data, as available. We will consider potential factors including: (1) The possible influence of differential non-response, especially among certain segments of Trump voters; (2) The role of likely-voter modeling, considering turnout, group size and vote estimates of different models overall and among groups; (3) Characteristics of those who declined to report their vote preference; (4) Group sizes, nationally and by geographical area, including any impacts of sampling and weighting procedures and (5) Potential interviewer effects.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Tracking the Election to Understand Trump's Win

Did the Tea Party Drive the Rise of Trump?

Gustavo Sanchez *Civis Analytics*

Geoff Bakken *Civis Analytics*

During Barack Obama's presidency, the Republican Party saw the development of two significant anti-establishment movements: the Tea Party and the presidential candidacy of Donald Trump. While the Tea Party contributed to victories of political newcomers over incumbents in Congressional races from 2010-2014, Trump and his supporters defeated a slate of more experienced politicians to seize the Republican nomination for president. While the preponderance of white, male, and less educated supporters in each movement suggests overlap (Zernike 2010; Thompson 2016), Trump's platform departed in key ways from the Tea Party's small government, free market orthodoxy. The question remains: to what extent were the Tea Party and the rise of Trump driven by similar members of the Republican Party coalition? This paper addresses this question by comparing supporters of the Tea Party following the 2014 midterm elections to supporters of Trump during the 2016 Republican presidential primary using survey data collected by data science firm Civis Analytics. The analysis compares the social and demographic characteristics of self-identified Tea Party supporters based on 15,097 telephone interviews conducted in January 2015 to self-identified Trump supporters based on 20,578 telephone interviews conducted between August 2015 and April 2016. Survey responses are linked to auxiliary data in Civis' proprietary consumer database to build regression models explaining support for the Tea Party and for Trump. In addition to building separate models for each set of interviews, the analysis utilizes machine learning techniques to calculate propensity scores for Tea Party support for all individuals in the consumer database, which has high coverage of the voting age population. This approach bridges the gap across the time periods and allows for an examination of the role of estimated Tea Party support in favoring Donald Trump among those surveyed in late 2015 and early 2016.

May 21st, 2017
8:30 AM - 10:00 AM
Concurrent Session J

Diversity: Attitudes and Measurements

Social media utterances and public opinion of income and racial inequality in the United States

Lucy Odigie Turley *The Opportunity Agenda*

Can social media data accurately predict public opinion? This question has been asked with increasing fervor in recent months, as the reliability of polling and survey research has come under increasing scrutiny. While a growing literature has pointed to the potential of social media for predicting election results (Franch, 2013; O'Connor, Balasubramanyan, Routledge, & Smith, 2010), few studies have focused on the role social media utterances can play in analyzing public perceptions of social justice issues, such as income and racial inequality. This research adapts the method of social media sentiment analysis put forth by Brandon Conner and colleagues (2008) and examines public attitudes towards income and racial inequality in the lead up to the 2016 presidential election. Analysis of social media data was conducted using Crimson Hexagon, a leading social media analytics software which provides access to proprietary data including, but not limited, to Twitter, Facebook, Instagram, Blogs, Forums/popular message board such as Reddit, and mainstream news article comments, reviews, and Youtube comments. Findings indicate that social media data offers reliable, temporal access to public opinion on a variety of social and political issues, as well as insights into the factors and information sources that currently influence public sentiment.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Collecting, Managing and Sharing Data - Using the Data Documentation Initiative (DDI) Standard across the Survey Research Lifecycle (Panel)

Collecting, Managing and Sharing Data - Using the Data Documentation Initiative (DDI) Standard across the Survey Research Lifecycle

Steven McEachern *Australian Data Archive*

Barry Radler

Daniel Gillman *US Bureau of Labor Statistics*

Barry Radler *University of Wisconsin*

Jon Johnson *Institute of Education (UCL)*

William Block *Cornell University*

Kathleen Weldon *Roper Center for Public Opinion Research, Cornell University*

Jared Lyle *DDI Alliance, University of Michigan*

The Data Documentation Initiative (DDI) is an international standard for describing statistical and social science data. This freely available international standard was first released in 2000 to enable the description of data that result from observational methods in the social, behavioral, economic, and health sciences. Today DDI is used by data producers, data managers, archives and national statistical institutes in over 85 countries to manage survey, administrative and other data. This panel is intended to provide an introduction to the Data Documentation Initiative, and how it can assist survey research and field agencies to Document, Discover, and Interoperate with their data. Papers will cover the content and usage of the DDI standard, using examples from across social, behavioural and health sciences. The papers in the session will introduce DDI, and illustrate how DDI is being used to produce instruments, data dictionaries, codebooks and other documentation and metadata required for managing survey data across the research lifecycle, including the collection, management, preservation and dissemination of survey and other data. The panel will commence with an introduction to the DDI standard, and then present 5 papers, covering the use of standards across the research lifecycle, including 1. Data collection and management: Lowering the barriers to capturing questionnaire metadata throughout the data lifecycle 2. Processing: Documenting Consumer Expenditure Survey Processing Using DDI 3. Discovery and Harmonisation: Harmonisation and discovery of longitudinal survey data using DDI in the UK CLOSER project 4. Dissemination: Using DDI for managing and disseminating public opinion research data at the Roper Center 5. Transparency: DDI and the AAPOR Transparency Initiative: Comparison and Future Directions

Lowering the barriers to capturing questionnaire metadata throughout the data lifecycle

Barry Radler

Sufficient metadata facilitates the discovery, use, and interpretation of survey research data and is necessary to maximize the value of collected data. Usually this information is appended to survey datasets or summarized in a codebook toward the end of the research data lifecycle, immediately prior to analysis, preservation, and/or storage. Yet this same metadata has often been captured, administered, or managed by different stakeholders numerous times at different points along the data lifecycle. Capturing research metadata earlier in the data lifecycle—from conceptualization and instrument development onwards—not only provides richer and more accurate information, but it also

introduces efficiencies to workflows by reducing redundancies and lowering costs. The Data Documentation Initiative (DDI) metadata standard can organize the means by which this metadata is tracked and managed at each stage of a survey's lifecycle. DDI has a comprehensive and richly structured XML schema that can pose challenges for clients, researchers, and project managers who aren't fluent in metadata standards or software. There may also be cultural and technical hurdles for these stakeholders to overcome when describing questions and assessments in the manner that DDI does. The core of the problem is how to achieve DDI buy-in early in the process so that metadata follows the data at every stage of its existence. This presentation proposes a means by which a simple Word template may serve as a bridge from conceptualization to instrument development to fielding. By providing a familiar interface that many survey clientele already use to develop questionnaires, this template could substantially lower the entry barriers to using DDI in survey research. The presentation will propose a mapping of the template fields to DDI schema elements and will elucidate how DDI can provide information on versioning, harmonization, and provenance in a standardized and machine-actionable manner.

Documenting Consumer Expenditure Survey Processing Using DDI

Dan Gillman

The Consumer Expenditure Survey (CE) is a US Bureau of Labor Statistics (BLS) program designed to measure how US families spend their money. The data are also used as input to the Consumer Price Index (CPI), issued monthly. CE management has identified the need to document all the processes used by BLS to transform CE data from collection to dissemination, DDI 3.2 was selected as the metadata framework for accomplishing this, and the suite of software from Colectica was selected to build a system. CE is conducted as 2 separate surveys, Interview and Diary. Then, the data are combined during processing. Through the process, data are packaged in 2 ways, one way for CE dissemination and the other for the CPI. Variables, code sets, questionnaire, and other features undergo changes in every odd numbered year. Estimates are created every 6 months for the previous year. Data are sent to CPI monthly. In addition, CE processing is divided into 4 phases, loosely defined as: Phase 1 – sample selection and collection; Phase 2 – initial edit subsystem; Phase 3 – estimation and edit subsystem, and data sent to CPI; and Phase 4 – final edits, tables, microdata. Management needs to be able to follow data through all the phases to track and manage the production process. Forward and backward paths need to be shown. So, a documentation system must be able to account for all these needs. To do this, the BLS is conducting a phased approach, starting simple and adding complexity as experience is gained. The incremental systems are designed to establish that DDI and the Colectica system are sufficiently sophisticated to handle the needs of CE. This paper will go into more detail about the particulars of the CE survey and describe results and plans for building a system.

Harmonisation and discovery of longitudinal survey data using DDI in the UK CLOSER project

Jon Johnson

The UK is home to the world's some of the longest-running longitudinal studies most notably the national birth cohorts starting in 1946, 1958, 1970 and 2000 and many other regional studies such as ALSPAC, the Southampton and Hertfordshire cohorts and the successor to the BHPS, Understanding Society. Cohort & Longitudinal Studies Enhancement Resources (CLOSER) aims to maximize their use, value, and impact both at home and abroad through the development of a rich metadata portal which has the broad aims of easing discovery, providing provenance of both questions and variables thereby encouraging comparison, harmonisation and reuse of questions. At the heart of the project is the assembly of a metadata repository encompassing the full range of information about the studies, from

study description to data collection methodology, questionnaire capture and data. Layered on top of this are a controlled vocabulary and other contextual information. The project has very specific challenges. The studies are based in six organisations geographically dispersed across England. Leadership of the studies crosses scientific domains (some primarily medical, some social science focused) and the longevity of the studies (starting as early as 1946) means that information available is inconsistently catalogued and held in various formats. Utilising DDI-Lifecycle, the project is creating a research discovery portal that will be interoperable with similar efforts and studies to enable discovery and comparison of questions and data. The project has primarily used off-the shelf software, combined with some lightweight tools has enabled mass ingest of metadata into a repository in a consistent and manageable process. The paper will focus on challenges of creating such a rich metadata repository, the lessons learnt and implications for reuse of questions in the development of future studies, in discovery for research and how such infrastructures can assist harmonisation and comparison within and across longitudinal studies.

Using DDI for managing and disseminating public opinion research data at the Roper Center

Bill Block, Kathleen Weldon

In 1957, polling leaders determined that a national archive of polling data, accessible to all practitioners of survey research, was necessary to move the field forward and open up new forms of data analysis based on the comparison and pooling of data from multiple sources. Since then, the Roper Center has held a unique position as the central repository for U.S. polling data. Roper has adopted DDI standards for data processing in order to: improve the efficiency of its internal data processing, provide more complete records for future researchers, and aid in the creation of discovery and analysis tools. By adopting DDI, Roper will greatly increase the granularity of its metadata, collecting and disseminating far more specific detail about polls and their methodology and saving that information as searchable, analyzable elements in its archive. This new metadata structure will enable new forms of analysis across the archival collection, allowing Roper to fulfill more fully the vision of its founding. In addition, the implementation of DDI standards at Roper will encourage compliance with the AAPOR Transparency Initiative for the many survey research organizations that archive with the Center, promoting greater sharing of information about methodology and practices across the industry. Roper will demonstrate a DDI-compliant data submission portal and explain how our use of DDI supports the mission of the Roper Center to preserve and disseminate public opinion data and holds the potential to bring improvements to the larger polling industry.

DDI and the AAPOR Transparency Initiative: Comparison and Future Directions

Steven McEachern, Jared Lyle

The AAPOR Transparency Initiative (TI) was formally established in 2014 following several years of development, and now has 40 approved member organisations. At the core of the initiative includes the reporting of survey research results according to the TI disclosure elements. This development has however occurred in isolation from another major initiative focussed on documentation for survey data management, the Data Documentation Initiative (DDI). This standard, developed under the auspices of the DDI Alliance, was established in 2000, with contributions by survey data collection organisations, statistical agencies, and social science data archives, each of which have a long history of supporting access to survey research across North America and around the world. This paper aims to reconcile these two initiatives to compare their intent and operation, and establish a possible path for coordinated use of the two frameworks. The paper begins with an examination the TI disclosure elements (<http://transparency.aapor.org/docs/TransparencyInitiativeDisclosureElementsMay2015.pdf>),

to outline the core requirements of TI reporting. These elements are then compared with the core content Of the DDI Codebook standard (<http://www.ddialliance.org/Specification/DDI-Codebook/2.5/>) to identify the relevant elements of DDI which could support each TI required element. Preliminary mapping of the two frameworks suggests that DDI could effectively be used to standardise this reporting of TI into a human and machine actionable framework for automated transparency reporting. The paper therefore concludes with suggestions for implementation of standard TI reporting requirements using the DDI metadata framework, and a discussion of future opportunities for possible integration between DDI and TI in survey data management

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Stemming the Tide of Nonresponse: Examining Study Design Characteristics that Influence Response Rates

Hello? It's You We're Looking For: Communicating with Panelists in the Digital Age

Meredith Czaplewski *Nielsen*

Erin Wittkowski *Nielsen*

With the proliferation of smart mobile devices in the past decade, the means researchers have at their disposal to contact panelists is at an all-time high. However, with this diversity comes the increasing ability for individuals to tailor the communications they receive by “opting out” of common methods of contact, potentially limiting researchers in their ability to connect with participants. At the same time, researchers increasingly run the risk of their communications getting lost amid the growing sea of other messages competing for the attention of panelists that don’t opt out. A panel’s success depends on the ability to communicate with its panelists in easy, clear, convenient, and reliable ways. While extensive research has been done on the effect that methods of contact have on response rate across survey modes, the effect that different communication channels have on researchers’ ability to connect to panel participants for day-to-day communication needs has been much less explored. In this paper, we focus on three main questions: 1) Who provides contact information through various communication methods, who gives us permission to use these methods to contact them, and which communication methods do participants prefer?; 2) Are we able to reach participants targeted for specific communications?; and if so, 3) How are communications through different channels received, interpreted, and used by participants to affect desired compliance behavior? To answer these questions, we utilize contact and permissions information from a Nielsen media panel and qualitative analysis from in-depth interviews to examine how demographic groups and language preferences differ in their communication preferences, permissions, and uses.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Internet and Mobile Data Collection: Compliance and Patterns of Use

What Can the General Social Survey's Web Version Teach us about Device and Platform Effects in Web Surveys?

Ipek Bilgen *NORC at the University of Chicago*

Michael J. Stern *NORC at the University of Chicago*

Tom W. Smith *NORC at the University of Chicago*

Although numerous studies have explored the effects of mode, platform, operating system, and device in web surveys, few have examined these influences in systematic experiments embedded in well-known national benchmark surveys such as the General Social Survey (GSS). In this study, NORC conducted a comparative study and designed first ever web version of the GSS survey using the items from the face-to-face GSS questionnaire. Our study compares items from the GSS web survey to the face-to-face GSS study, which serves as the benchmark for our analysis. The sample for the 20 minute web survey included 10,000 U.S. household addresses to which we mailed a series of survey invitation letters that included a link to the web survey, a \$5 pre-completion incentive, and a mode-choice postcard offering an SAQ booklet version of the questionnaire if sample members are not able to or prefer not to complete the survey via web instrument. Additionally, as part of follow-up efforts, we telephoned non-responders in an effort to drive them to the web if they have internet access. This paper covers: 1) the impact of the questionnaire length and wording complexity of selected items among mode, device, and platform; 2) compares responses from our study with the face-to-face GSS (i.e. benchmark comparison) among mode, device, and platform; and 3) examines the impact of internet platform/access and geographic clustering on coverage via comparison to the face-to-face GSS benchmark study. Our results aim to shed light into mode, device, platform, and operating system effects on measurement and coverage, and will enable us to tailor our data collection strategies to drive more households to the less-expensive web mode.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Stemming the Tide of Nonresponse: Examining Study Design Characteristics that Influence Response Rates

An examination of seasonal response rates during a year-long mail data collection using an ABS frame

Eric Jodts *Westat*

Sharon Lohr *Westat*

Most surveys are limited to a shortened time frame for reasons of expediency and timeliness. The few surveys with longer field periods often release all sample up front, while comparisons of independent surveys are often subject to confounding variables such as differences in protocol, content, sponsors, materials, instruments and so forth. Such factors hinder a straightforward evaluation of seasonal response rates. Our study gauged respondent annoyance on a number of neighborhood environmental factors. Since these attitudes can vary depending upon season our design required a 12-month field period, with periodic sample release, to capture potential seasonal variation. Our population of interest were households exposed to a certain level of noise around selected U.S. airports. Therefore, an address-based sample (ABS) was most appropriate. Using an ABS frame we released sample in 6 waves, each two months apart to provide a full year of data. The mail protocol for each wave took 6 weeks from start to finish providing us with daily responses over the data collection year. Here we present the results of our study showing the response rate impact of season on an ABS mail survey where all other variables remained constant. Our findings were not always in alignment with conventional wisdom. We will provide considerations for timing your mail data collection or adjusting sample to address potential fluctuations.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Stemming the Tide of Nonresponse: Examining Study Design Characteristics that Influence Response Rates

Adding Cell Phones to the American Community Survey Telephone Operation

David Raglin *U.S. Census Bureau*

The US Census Bureau's American Community Survey (ACS) uses four mode of data collection: Internet, mail, computer-assisted telephone (CATI), and computer-assisted personal visit (CAPI). In recent years, the CATI operation has seen significant declines in completion rates and increased costs. The CATI operation uses telephone numbers provided by vendors matched to ACS sample addresses. Currently, the ACS only calls landline telephone numbers. Unfortunately, about 40 percent of the telephone numbers do not reach the sample household. To improve the CATI operation the ACS is exploring using cell phone numbers in addition to landline numbers. To test the feasibility of using cell phone numbers to contact addresses eligible for CATI, we designed a experiment using the January 2017 ACS CATI operation. The panel was divided into two treatments: a control treatment where the best landline was called (the current ACS method), and a test treatment where the best landline or cell phone was called. The research examines the outcomes of calls for each treatment, including the completion rate and the percentage of numbers that reach the correct address. Additionally, it looks at other measures such as the time of day for completed calls, and compares these measures by treatment. Finally, the research investigates differences between treatments in the number of late Internet and mail returns spurred by the CATI calls.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Apt. 3, We Pick You. Resident 1, We Pick You Too: Improving ABS Surveys and Household Selection Methods

Comparison of three methods to select a respondent for household online surveys using mailed invitations

Keven Bosa *Statistics Canada*

François Gagnon *Statistics Canada*

Pierre Caron *Statistics Canada*

In populations with widespread access to and use of the internet, online self-reporting surveys can be a cost-efficient alternative to traditional modes of collection. Consider a sampling frame of households with mailing addresses. A random sample of households are sent paper invitations in the mail to complete an online survey. Within a sampled household, we want to randomly select an occupant to complete the survey. For this with-in household sampling, households are often asked to provide a roster of the eligible occupants. In a multi-person household, if an occupant provides a roster of the household online, a different occupant may be randomly selected to complete the survey. In this scenario, one occupant has invested time to go online and there is a risk the survey will not be handed over to the selected occupant and/or the survey will not be completed at all. Many statistical agencies have observed a decrease of the response rate in household surveys and this hand-off issue will exacerbate this decrease. In an attempt to increase the response rate, two alternative probabilistic selection methods were tested in a pilot survey: the last birthday method and a new variation of the age-order method. These methods avoid hand-off during online activity by selecting a respondent from the household using instructions on the paper invitation. In the pilot survey, the two alternative selection methods led to higher response rates compared to the traditional roster method. During this presentation, the three selection methods will be compared in terms of their response rates and selection inaccuracy rates.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Smarter Surveys for Smartphones: Optimizing Data Collection Using Mobile Devices

Optimizing a Government Household Survey for Mobile Devices

Jennifer Anderson McNulty *Westat*

Jocelyn Newsome *Westat*

Kerry Levin *Westat*

Brenda Schafer *Internal Revenue Service*

Patrick Langetieg *Internal Revenue Service*

Saurabh Datta *Internal Revenue Service*

Smartphone and tablet ownership has skyrocketed, with 68% of individuals owning a smartphone and 45% owning a tablet in 2015 (Pew Research Center, 2015). Increasingly, individuals opt to take surveys on mobile devices, with as many as 30% of respondents completing web surveys on a mobile device, although exact rates vary by population and topic (Peterson, 2012; De Bruijne, 2014). Given this trend, we incorporated a mobile version into the latest administration of our IRS Individual Taxpayer Burden (ITB) survey. The IRS ITB survey is an annual survey sent to 20,000 individuals that measures the time and money taxpayers spend complying with tax law. The survey comprises of two critical items that ask about time and money, along with 24 other items that provide context by asking about filing activities. In order to make the survey mobile-friendly, it was shortened to five items: the two items of highest research interest to the IRS—the time and money spent on the tax return—and three additional items that measured stress associated with complying with federal income tax filing requirements. The mobile survey was sent to a sample of 2,000 respondents and offered in English and Spanish. Respondents were able to access the survey by scanning a QR code or by typing in the URL provided in the instruction card they received in the mail. In this paper, we will examine the mobile survey's response rates, particularly from populations that have been historically underrepresented in the ITB survey (e.g., younger adults, low income respondents, and parents with young children). We will also look at which devices (i.e., laptop, tablet, or smartphone) respondents utilized when completing the survey. Finally, we will discuss respondents' preferred method for entering the survey (QR code versus URL).

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Basket of Deplorables: Race, Gender, Age and the Vote

Millennials and the 2016 Election: How race and ethnicity shaped young adults' experiences and beliefs

David Sterrett *NORC at The University of Chicago*

Liz Kantor *NORC at the University of Chicago*

Jennifer Benz *NORC at the University of Chicago*

Trevor Thompson *NORC at the University of Chicago*

Emily Alvarez *NORC at the University of Chicago*

Most survey research during the 2016 presidential campaign treated Millennials as a monolithic group and did little to examine racial and ethnic differences. Yet, Millennials are the largest and most ethnically diverse generation of Americans. About 19 percent of Millennials identify as Latino or Hispanic, 13 percent as Black or African American, and 6 percent as Asian American. To better understand the political attitudes and experiences of Millennials across races and ethnicity, the Black Youth Project at the University of Chicago with The Associated Press-NORC Center for Public Affairs Research developed GenForward, a first of its kind monthly survey. GenForward pays special attention to how race and ethnicity shape young adults' experiences and beliefs about the world. Each month, GenForward features a nationally representative survey of more than 1,750 adults age 18-30, including at least 500 Latino/as, 500 African Americans, 500 whites, and 250 Asian Americans. In this paper, we focus on racial and ethnic differences in young adults' attitudes and experiences in the months leading up to the November 2016 election. The findings from six different surveys highlight stark racial and ethnic differences in Millennials' beliefs about racism, policing, immigration, economic issues, and evaluations of the presidential candidates. Understanding the differing attitudes of young African Americans, Latino/as, Asian Americans, and whites leading up to the 2016 election provide important insights into Millennials' priorities and expectations for the new administration.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Basket of Deplorables: Race, Gender, Age and the Vote

Latinos in the 2016 Election: Was there a Trump effect?

Mark Hugo Lopez *Pew Research Center*

Ana Gonzalez-Barrera *Pew Research Center*

Gustavo Lopez *Pew Research Center - Washington, DC*

The 2016 election placed a great focus on the record 27.3 million Hispanic eligible voters and their political values and attitudes. At the beginning of his campaign, Donald Trump made provocative comments about Mexican and Muslim immigrants and Hispanics in general. The issue of immigration was also a focal point of Trump's campaign. The 2016 presidential election drew strong responses among the Latino community overall, but did Trump's comments motivate Latino voters more than in past election years? This analysis will be based on a number of Pew Research Center surveys done prior to the election, media exit polls done on the night of the election and the latest data from BLS and Census Bureau on civic engagement. Findings highlight that although a majority of Latino voters were engaged in the election and had discussed Trump's comments about Hispanics; this does not appear to have had a strong effect on voter engagement or voter turnout. A month before the election, 69% of Latino registered voters said they were absolutely certain they would vote on this election, compared with 77% who said so in 2012. It is likely that in 2016 the number of Hispanic voters reached a record number, but voter turnout rate might have remained at levels similar to past elections or increased slightly, with only about half of those eligible to vote casting a ballot.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Apt. 3, We Pick You. Resident 1, We Pick You Too: Improving ABS Surveys and Household Selection Methods

The Effects of Cell Phone Recruitment of Nonresponders in a Mixed-Mode ABS Study

Claudette Brazle *Simmons Research*

Gerry Dirks *Simmons Research*

Paola Pino *Simmons Research*

Paul J Lavrakas *Independent Consultant*

Each year, a larger proportion of survey research is being conducted using mixed-mode designs, where more than one mode is being used to sample, recruit, and/or gather data. Nowadays, many of these surveys use only an address-based frame/sample. As a first form of recruitment, sampled households in these studies often are mailed materials to invite/encourage their participation by either returning a hard copy of the questionnaire or going to a web site to complete a CAWI version. After multiple attempts to gain cooperation using the mail mode to reach the household, some surveys are trying to gain cooperation by dialing phone numbers that have been matched to the nonresponding addresses, and then gather data via telephone. Heretofore, these phone follow-ups have been limited to landline dialing. However, recent advances have been made in matching cell phone numbers to addresses. In 2016, the Simmons Research company conducted an experiment, within its National Consumer Survey, whereby a random national sample of households that was sampled via an ABS frame, but who had not responded to multiple mail contacts, but for whom a cell phone number could be matched to the address, were followed-up using manual dialing of matching cell phone numbers. It was found that by adding a nonresponse follow-up stage to the mixed-mode design using these matched cell phone numbers for recruitment and data collection, several benefits were achieved. These included raising the response rates (1) overall to the first stage of this two-stage survey, (2) among sampled addresses from a previous wave that had ended as nonresponders, (3) among addresses with Hispanic residents, and (4) among Cell Phone Only households. In addition, and surprisingly, manual dialing of matched cell phone numbers was found to be twice as productive in gaining completions compared to predictive dialing of matched landline numbers.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Internet and Mobile Data Collection: Compliance and Patterns of Use
Examining Potential Sources of Nonresponse to Mobile Data Collection with New Technologies in a Probability Household Panel

Alexander Wenz *University of Essex*

Annette Jäckle *University of Essex*

Mick Couper *University of Michigan*

The emergence of mobile technologies, including smartphones and tablets, offers new possibilities for survey research. While conventional web questionnaires can be completed through a web browser, additional data can be collected on these devices using their integrated features, such as GPS, cameras, and environmental sensors. When conducting studies that involve mobile data collection, however, researchers must keep in mind that not all parts of their target population might have access to mobile devices, or might be willing to use their device for data collection purposes. This paper investigates the feasibility of using new mobile technologies to collect data from the general population, and aims to identify potential sources of nonresponse. We asked 2,176 members of the UK Household Longitudinal Study Innovation Panel, a nationally representative household panel study in the United Kingdom, about their device usage behaviour and about attitudes towards sharing information using mobile devices. First, we will gauge what proportion of the general population has access to mobile technologies and mobile Internet, and examine biases in the population with regard to device ownership and Internet access, considering socio-demographic characteristics. Second, we will investigate which activities mobile devices are currently being used for, and assess self-reported skills of using a mobile device. Third, we will examine the hypothetical willingness of our sample to participate in various data collection tasks using their mobile device, such as sharing their GPS position, connecting their device to other electronic devices via Bluetooth, or providing device use data. Fourth, we will investigate data security concerns of the population when providing information with mobile technologies compared to questionnaire-based methods of data collection, including face-to-face, telephone, mail, and web surveys.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

How Good Is Cheap? Evaluating the Quality of Estimates Derived from Non-Probability Samples

Assessing Child Vaccine Hesitancy using Mobile Panels

John Boyle *ICF International*

Lewis Berman *ICF International*

James Dayton *ICF International*

Deidre Middleton *ICF International*

Alex Coleo *mFour*

The Healthy People 2020 Objectives established a target of at least 90% of 2-year-old children fully vaccinated with most recommended vaccines. This is a critical goal in establishing “herd immunity” in the general population. However, traditional survey methods are expensive and slow for monitoring immunization rates in low incidence populations like young children. This study was designed to test alternative sampling and data collection methodologies for assessing childhood immunization more efficiently than traditional methods. A national sample of adults from a mobile panel, who had children in the household, were invited to participate in an app-enabled survey on their smartphone. More than 1,000 respondents with children aged 19 to 35 months completed an interview in a seven day field period. The interview, patterned on the National Immunization Survey (NIS), took approximately 10 minutes. The characteristics of the parents and children in the survey sample were generally comparable to those of the NIS. The web panel approach allowed us to explore “vaccine hesitancy” to childhood immunization in this national sample of parents of young children. Due to concerns about interview length, the NIS has not explored the reasons for refusal or delay in childhood immunization in nearly a decade. This survey examines current attitudes of parents of young children toward childhood immunization, and how these attitudes are related to complete and up-to-date immunization rates in American children.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

The Climate Change "Conspiracy"

Religious Leaders and Public Opinion on Climate Change: Priming Pope Francis Heightens Moral Perceptions of the Issue

Jonathon P. Schuldt *Cornell University*

Adam R. Pearson *Pomona College*

Rainer Romero-Canyas *Environmental Defense Fund*

Dylan Larson-Konar *Environmental Defense Fund*

In his recent encyclical letter *Laudato Si'*: On Care for Our Common Home, Pope Francis argued that taking action to address climate change is a moral imperative. However, social science research suggests a complex relationship between religious concepts and environmental attitudes, raising the question of what influence the pope's position may have on public opinion regarding this polarizing issue. In a national probability survey experiment of U.S. adults ($n=1,212$), we find that brief exposure to Pope Francis influenced the climate-related beliefs of broad segments of the U.S. public: it increased perceptions of climate change as a moral issue for the overall sample (and among Republicans in particular) and increased felt personal responsibility for contributing to climate change and its mitigation (among Democrats). Moreover, prior awareness of the pope's views on climate change mattered, such that those who indicated greater awareness of the pope's position showed stronger treatment effects, consistent with a priming account of these effects. Results complement recent correlational findings and offer further evidence of the Vatican's influence on climate change public opinion.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Stemming the Tide of Nonresponse: Examining Study Design Characteristics that Influence Response Rates

Which Interviewer Training Characteristics Improve Survey Data Quality? A Meta-Analysis

Jessica Wengrzik *GESIS - Leibniz Institute for the Social Sciences*

Michael Bosnjak *GESIS - Leibniz Institute for the Social Sciences*

Relevance & Research Question: The aim of this meta-analysis is to (1) explore whether interviewer training can significantly improve interviewee cooperation, the data and paradata quality and (2) which approaches used to train survey interviewers have been most successful? Herewith we focus on the role of interviewer training to improve cooperation rates, interviewer error rates, unit nonresponse rates, socially desired behaviour as well as interviewer reliability and accuracy of respondent rating. Telephone and face-to-face surveys claimed in 2016 more than half of all surveys worldwide (ESOMAR, 2016). Those surveys served in many cases as political and economic decision-making basis. However many authors showed that there is a strong link between interviewer qualification and data quality (Billiet 1988; Dahlhamer 2010; Olson 2007). An adequate way to qualify interviewers is via interviewer training. Many huge survey projects as PIAAC (PIAAC, 2014) or the ESS (Loosveldt et al., 2014) expect well trained interviewers and survey institutes provide trained interviewers. But what characteristics a successful training involves is still a “black box”. Interviewer training approaches are heterogenous in method, content and lengths. This meta-analysis aims to answer the question which interviewer training characteristics make a training successful to advance advice for optimizing interviewer trainings to increase data and paradata quality? Concrete research questions are: How long should a successful interviewer training take? How much can Anti-Refusal-Training improve the interviewee cooperation? Which role play practice and feedback versus instruction only? Is online training as successful as training on-site? Does it make a difference training face-to-face or telephone interviewer? **Methods and Data:** We identified almost 40 interviewer training experiments for this meta-analysis. Our effect size is data and paradata quality operationalized through completion rate (>20 studies), interviewer errors (>30 studies), accuracy and halo (>80 studies), social desired behaviour, interviewer retention and item nonresponse.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

How Good Is Cheap? Evaluating the Quality of Estimates Derived from Non-Probability Samples

What Can We Infer from a Non-Probability Sample of People with Diabetes? Developing Adjustments to Results from a Commercial Access Panel.

Thomas M Guterbock *University of Virginia, Center for Survey Research*

Kara ShanerFitzgibbon *University of Virginia Center for Survey Research*

Hyojung Kang *University of Virginia Dept. of Systems and Information Engineering*

Jennifer M Lobo *University of Virginia Dept. of Public Health Sciences*

Ishan C Williams *University of Virginia School of Nursing*

Min-Woong Sohn *University of Virginia Dept. of Public Health Sciences*

We developed a survey to gather information quickly from U.S. adults aged 40 or older who have diabetes. The purpose was to learn whether these health consumers are aware of certain off-label benefits of statin drugs, including reduced risk of lower-limb amputation, and to measure willingness to take statins for their possible off-label benefits, even in the absence of high-cholesterol problems. We gathered on-line responses to a detailed, 15-minute questionnaire from 339 qualified respondents reached through SSI's access panel, which is recruited by non-probability methods. Our survey included several questions on diabetes management and self-care and guideline-recommended care utilization drawn from BRFSS—a large, government-sponsored survey that uses probability methods of sampling. Merging our data set with this reference survey, we use binary regression to compute pseudo-inclusion probabilities for sample cases (Valliant & Dever 2011), based on demographics, health status variables, and reports of some self-care practices. We then apply propensity score adjustments to our sample and test the representativeness of the adjusted sample by comparing adjusted results for other self-care variables to those from the reference sample. A second adjustment strategy is to treat the reference sample as a superpopulation and calibrate our sample, using raking, to control totals drawn from the reference sample. A third strategy is to first apply the propensity weights, then calibrate to the superpopulation control totals. For each of these strategies, we examine the accuracy of our sample estimates for variables existing in the reference sample, and consider the impact of adjustment on our estimates of the new variables of interest. This work evaluates the utility of internet access panels for diabetes research, and also serves as a case study in the extent to which a non-probability sample of a relatively rare population can be adjusted to yield credible inferences about that population.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Basket of Deplorables: Race, Gender, Age and the Vote

Polling Millennials in 2016

Joshua J Dyck *UMass Lowell*

John Cluverius *UMass Lowell*

As coverage and response bias issues become more difficult to overcome in traditional telephone surveys, this reality is amplified for young voters, who are the least likely to respond to telephone surveys. The youngest voters in samples are often those who have the highest associated sample weights in order to achieve representativeness. Inferences about young Americans and their political preferences, then, are poorly understood and subject to large estimation errors. This paper explores several different options for weighting an online convenience sample of young Americans (aged 18-35) recruited in October of 2016. Online samples are ideal for studying the behavior of young Americans because the presence of coverage issues is attenuated. Our weighting approach include demographic weighting, demographic weighting with political variables (targets for voter registration and party identification), as well as a matching procedure. We examine how various weighting schemes affect what we know about how Millennials voted in the 2016 Presidential election. We find that the exit polls provide poor estimates of the Millennial vote, particularly among non-white voters, who make up over 40% of 18-35 year olds in the United States.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Basket of Deplorables: Race, Gender, Age and the Vote

Unpacking the Women's Vote

Jennifer Su *Princeton Survey Research Associates International*

Cary Funk *Pew Research Center*

Maureen Michaels *Michaels Opinion Research, Inc.*

Hannah Hartig *Penn Program on Opinion Research and Election Studies*

Stephanie Perry *NBC Universal*

Hillary Clinton was the first woman to lead a major party's bid for the presidency and yet, like so many forecasts before this, 2016 was no 'Year of the Woman.' This paper will use the 2016 (NBC News Exit Polls) to unpack the factions within the women's vote, both nationally and in key battleground states, that helped Donald Trump win. To do so, we will compare the women's vote from past presidential elections with the 2016 vote, paying particular attention to white conservative women supporting Trump. We will also look at the divisions within the Democratic Party, especially by generation, that supported a more liberal agenda in the primaries and how these women voted in the general election. We compare these findings to a handful of U.S. Senate races with women candidates to better understand the conditions when women running for office get a boost at the polls. We will discuss the implications for women presidential candidates in the years to come.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Basket of Deplorables: Race, Gender, Age and the Vote

Emasculation and 2016: Gender Role Threat, Attitudes and the Vote

Dan Cassino *Fairleigh Dickinson Univ*

Peter Woolley *Fairleigh Dickinson University*

It seems clear that gender remains a potent force in American politics, and that gender role threat, in particular, plays a significant role in men's attitudes, issue positions and vote choices. For both men and women, gender role threat results from factors in the household, community or society that lead to the perception that the individual's gender role is being called into question. For instance, men who have based their masculinity on their role as breadwinners may face gender role threat if their spouse begins to earn more than they do. Such gender role threat can lead to dramatic shifts in attitudes and issue positions that are perceived to be related to gender, such as abortion, gay rights and gun ownership. These issues have become especially fraught in the wake of the 2016 US Presidential election. This paper presents evidence that while individual gender role threat has been on the decline, perceptions of societal gender role threat have been on the increase, and this perceived societal threat pushed many voters against Democratic nominee Hillary Clinton. In addition, the evidence suggests that many women, especially in older age cohorts, are also using political behavior and attitudes as a way to compensate for threat to their gender roles.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

How Good Is Cheap? Evaluating the Quality of Estimates Derived from Non-Probability Samples

Fair Market Rent Estimation – Using Auxiliary Data and Non-Probability Samples to Calculate Fair Market Rents

Thomas Brassell *ICF International*
Randal ZuWallack *ICF International*
Brian Orleans *ICF International*

Annually, the Department of Housing and Urban Development (HUD) uses the American Community Survey (ACS) data to calculate rental subsidy levels (defined in most cases as the 40th percentile market rent) for metropolitan statistical areas (MSAs) and counties across the United States. Public Housing Authorities (PHAs) may contest these FMRs by conducting their own local market rent surveys. Recently, ICF conducted multiple FMR studies in various regions and calculated a 40th percentile rent that was 11% - 30% higher than that calculated using ACS data. The underestimation of the subsidy level warrants the need for further exploration of alternate methods for calculating FMRs. The current study explores both the use of auxiliary data and non-probability samples to estimate FMRs. For the auxiliary data comparison, we will use rent data collected by PHAs, and Zillow, respectively, to calculate the 40th percentile rent for large and small metropolitan regions. We will also compare the rental cost trends of these data sources and the ACS. We hypothesize that both rental data sources will produce higher 40th percentile rent estimates than ACS data. In addition, we hypothesize the rental cost trends produced by these data sources will demonstrate significantly higher increase in market rents than that produced by the ACS data. To assess the efficacy of a non-probability study, we will compare the 40th percentile rents calculated from a non-probability sample, and a probability address-based sample. Although research has found limited support for the generalizability of non-probability samples (Yeager, Krosnick, Chang, Javitz, Levendusky, Simpser, & Wang, 2011), we hypothesize that due to the more objective nature of the measure (specifically rent and utilities paid monthly), non-probability samples will produce similar FMR estimates, warranting a revisit to the currently accepted HUD FMR Survey methodology.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

The Climate Change "Conspiracy"

The Primacy of Politics in Views on Climate Change

Brian Kennedy *Pew Research Center*

The election of President Donald Trump figures to result in major changes in climate and energy policy, making it particularly important to understand how the American public sees these issues. Drawing on a recent Pew Research Center survey, we find that political polarization on climate and energy issues extends far beyond whether climate change is occurring and whether humans are playing a role. There are wide political divides across every dimension of the debate, from how the public trusts experts on climate change, to opinions about possible solutions, such as restrictions on power plant emissions, to views on possible consequences, like more severe storms. Even after controlling for demographics and competing explanations in statistical modeling, political party identification and political ideology are the two dominant factors shaping views. Moreover, to the extent science knowledge influences views about climate change and judgements about climate scientists, it does so among Democrats, but not Republicans. For example, in statistical modeling, Democrats with high science knowledge are more likely to believe climate change is the result of human activity and trust climate scientists, but science knowledge has no effect among Republicans. This leaves Americans with high science knowledge the most divided by politics in their views on climate change.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

The Climate Change "Conspiracy"

Trump Voters and Global Warming

Seth A. Rosenthal *Yale Program on Climate Change Communication*

Anthony Leiserowitz *Yale Program on Climate Change Communication*

Edward Maibach *George Mason University Center for Climate Change Communication*

Connie Roser-Renouf *George Mason University Center for Climate Change Communication*

Matthew Cutler *Yale Program on Climate Change Communication*

Geoff Feinberg *Yale Program on Climate Change Communication*

In 2012, Donald Trump notoriously tweeted that “the concept of global warming was created by and for the Chinese to make U.S. manufacturing non-competitive,” one of numerous Trump tweets either questioning the existence of global warming or challenging policies aimed at addressing it. But do the voters who propelled him to office share his dismissive views? Results from two representative national surveys—one conducted in March 2016, during the primary process, and a second conducted in November 2016, just after the General Election (still in the field during the writing of this abstract)--address this question. The March survey indicates that a considerable number of Trump’s strongest supporters do not fully share his attitude. Although they do not prioritize global warming as an election issue (it ranked last in importance of 23 issues we examined), they are more mindful of global warming’s existence, causes, consequences, and remedies than might be predicted. For instance, more than half of Trump’s primary supporters (56%) think global warming is happening and one in three (35%) are “very” or “somewhat worried” about it. They also support policies that are crucial for mitigating global warming, such as funding research into renewable energy (76% support), regulating CO₂ as a pollutant (62%), placing strict limits on emissions from existing coal-fired power plants (the backbone of the EPA’s Clean Power Plan--50%), and taking action to reduce global warming even if other countries do not (49%). However, they also favor fossil-fuel-friendly policies such as the expansion of offshore oil drilling (76%). While Trump’s statements and transitional appointments do not bode well for continued U.S. action on global warming, the opinions of Trump’s supporters regarding global warming’s consequences and solutions offer a notable contrast.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Smarter Surveys for Smartphones: Optimizing Data Collection Using Mobile Devices

Practical guidelines for developing a smartphone-based survey instrument

Jakob Ohme *University of Southern Denmark, Center for Journalism*

Claes de Vreese *University of Amsterdam, ASCoR*

Erik Albaek *University of Southern Denmark, Center for Journalism*

With the increasing popularity of mobile devices a new track of research developed, using mobile devices in survey research (Link, Murphy, Schober, Buskirk, Childs & Tesfaye, 2014). But the novelty of such smartphone-based tools makes it important to gather empirical evidence on designs of such surveys. Advantages of mobile measurements are, among others, timely answers, shorter but more frequent measurements and subsequently less pressure on respondents' memory. Obstacles, however, are mostly seen in a resource-intensive development and uncertain cooperation rates of respondents. This study presents the results of a test study conducted to identify the best set-up for a smartphone-based survey. The final purpose of this survey was to develop a mobile media measurement, using a smartphone-based diary approach with multiple subsequent measurements. Such an approach is seen as valuable to receive a more comprehensive assessment of peoples' media use, especially towards the background of converging media channels and short-term exposure via social media platforms. We base our analysis on a random sample of Danish citizens (n=196) who participated in mobile surveys during three subsequent days of investigation. First, we investigated whether survey invitations sent via an app-based push message or via text message result in different response rates. Second, we tested for differences between the two platforms regarding response time lag. Third, we tested whether send-out time (evening vs. morning) affects the number of respondents taking part in the survey. Our findings show that choice of platform as well as send-out time influence response behavior. Based on the good performance of the survey across platforms, we suggest that in certain contexts mobile-only surveys are a viable option compared to more traditional surveys.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Smarter Surveys for Smartphones: Optimizing Data Collection Using Mobile Devices

Willingness to participate in passive mobile data collection

Florian Keusch *University of Mannheim*

Christopher Antoun *U.S. Census Bureau*

Mick Couper *University of Michigan*

Frauke Kreuter *Univ. of Maryland, JPSM*

Bella Struminskaya *Utrecht University*

The rising penetration of smartphones now gives researchers the chance to collect data from smartphone users – such as geolocation, online behavior and browser history, app usage – through passive mobile data collection via apps. Compared to surveys that rely on self-reported data, passive data collection has the potential to provide richer data (because it is almost constantly collected), to decrease respondent burden (because fewer survey questions need to be asked), and to reduce measurement error (because of less forgetting and social desirability). However, to collect passive data from smartphones, participants need to agree to download a research app that tracks their behavior and location over a longer period of time. This leads to concerns about nonresponse and consent as well as privacy concerns. In the current study, we assess the circumstances under which smartphone users are willing to participate in passive mobile data collection. The study comprises two waves with approximately 1,800 respondents in a German nonprobability online panel. In wave 1, we screen for smartphone ownership and ask questions about topics such as smartphone and app use, attitudes toward survey organizations and passive mobile data collection, and general privacy concerns. Two weeks later, in wave 2 we show respondents a set of vignettes that describe hypothetical studies where data are automatically collected by a research app on a participant's smartphone. The vignettes vary the levels of several dimensions of the hypothetical study (e.g., sponsor, topic, study length, incentives), and respondents are asked to rate their willingness to participate in such a study. The study contributes to the literature on factors influencing participation decision in research in general and for the new application of mobile passive data collection specifically.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Apt. 3, We Pick You. Resident 1, We Pick You Too: Improving ABS Surveys and Household Selection Methods

Household Selection Using Birthday Methods in Self-Administered Surveys

Naomi Freedner *ICF International*

John Boyle *ICF International*

Robert Tortora *ICF International*

Kisha Bailly *ICF*

The (most recent/next) birthday method has been widely used in telephone surveys as an efficient method of respondent selection in households with multiple eligible respondents. It is a systematic, although non-random, household selection procedure. The slight age bias in either birthday method can be corrected by alternating or randomizing “most recent” and “next” protocol. This selection procedure has also been applied in self-administered surveys using address based sampling. This is more problematic for researchers because there is no interviewer to explain the selection process or record the outcome. In one important early study, Battaglia found that often the selection of the respondent with the (most recent/next) birthday was not accurate. As a result of a coding error in the implementation of a national ABS survey, a sample of 6000 respondents who initially responded by web were subsequently sent a request to complete a paper questionnaire. Approximately 1,000 respondents at the same address selected by the birthday method completed the same interview approximately four to eight weeks apart. Based on gender and age, these respondents can be classified as the same or different. This presentation examines the characteristics of households and individuals to determine the extent to which the birthday method obtains the same or different individuals in the same households. It also examines the extent to which the same or alternate selection of the designated respondent in the same household affects key survey estimators.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

How Good Is Cheap? Evaluating the Quality of Estimates Derived from Non-Probability Samples

Using online panel surveys to estimate population-level health statistics

Reanne Townsend *Westat*

Rosalynn Yang *University of Maryland*

J. Michael Brick *Westat*

Terisa Davis *Westat*

Richard P. Moser *National Cancer Institute, NIH*

Gordon Willis *National Cancer Institute, NIH*

Stephanie Fowler *National Cancer Institute, NIH*

David Berrigan *National Cancer Institute, NIH*

Dana Wolff-Hughes *Behavioral and Social Science Research in the Office of the Director, NIH*

Kelly Blake *National Cancer Institute, NIH*

Despite the increasing use of non-probability samples within survey research and previous studies comparing results between probability and non-probability samples, it is not yet clear under which circumstances appropriately weighted non-probability samples can yield valid inferences of population-level statistics. Last year we presented preliminary results from a study exploring the differences in health-related outcomes between a national probability mail survey and 2 non-probability web surveys. Several key estimates, (such as the proportion of respondents who have never smoked) were similar across methodologies, but others (especially those associated with use of the internet) had large discrepancies that were not corrected by propensity weighting. The current study extends this analysis, using additional data, into a more detailed and systematic review of the differences in population-level estimates of health outcomes between probability and non-probability sampling designs. We will compare results from the Health Information National Trends Study (HINTS), a national probability-based mail survey, to online panel surveys from four different vendors, using a range of sampling methods including probability, quota-based, and convenience. First, we compare estimates from different sampling methodologies to eight national benchmark estimates, and examine whether differences from benchmarks are vary between demographic groups. Second, we reproduce key associations from the health literature (such as the relationship between BMI and physical activity), and determine whether data from online and non-probability samples yield similar inferences. Together, these analyses will contribute to clarifying conditions under which online panel data can be used for valid inferences concerning health information knowledge and health behaviors.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Survey Nonresponse in 2016 Election Polling

Not Getting It Wrong Again: Leveraging a Bayesian Approach to Reduce Non-Response Bias in 2016 Pre-Election Polls

Chris Jackson *Ipsos*

Neale El-Dash *Ipsos*

Joe Zappa *Ipsos*

One of the many shocks surrounding the 2016 Presidential Election is, the accuracy, or lack thereof, of pre-election polling. Prior to November 8, the large majority of polls showed Hillary Clinton with a clear lead both nationally and in key battleground states. Our new national polling tool – the Reuters/Ipsos States of the Nation (SOTN) poll -- was part of this preponderance of polling showing the election close, but in Clinton's favor. However, Clinton lost in several key swing states leading Donald Trump to win the presidency through the electoral college. Our post-election analysis indicates that pre-election surveys, even those with tens of thousands of interviews, lack the sample size to accurately represent all potential groups in the electorate. This paper shows how the use of a Bayesian approach allows better estimates of population voting intention than more traditional non-probability or probability based election polling. We do this by comparing the results of our more standard polling approach of the Reuters/Ipsos SOTN polling against a combination of SOTN data with Census statistics. The SOTN poll, a rolling national sample of 15,000+ national non-probability online interviews provides information on the sentiment of voters while the Census data allows us to more accurately estimate populations at a granular level. The combination of the two data sources allow election forecasts more accurate than what was produced pre-election.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Hello! I am Your Interviewer and I May or May Not Affect Your Data Quality

Interviewer Attitudes Towards Respondent Persuasion: The Impact on Production

Lauren M Machingo *RTI International*

Stephanie Parker *RTI International*

Christina Touarti *RTI International*

Gretchen McHenry *RTI International*

Grace Medley *SAMHSA*

Barbara Forsyth *SAMHSA*

The National Survey on Drug Use and Health (NSDUH) conducted a series of short surveys with active interviewers in early 2016 to better understand field interviewer attitudes in the field and the impact those attitudes may have on respondent cooperation. Previous literature shows that interviewer attitudes and characteristics impact survey cooperation in a variety of ways (Durrant, et al., 2010). Across the survey series given to NSDUH interviewers, topics included overall job satisfaction, supervisor management and project expectations, persuading respondents, and communication and support. Questions related to respondent persuasion were developed from a similar survey conducted by the National Centre for Social Research (NatCen), which examined the role of interviewers' experience, attitudes, personality traits and inter-personal skills in determining survey cooperation. In their research, NatCen found that interviewers who "are more positive about the justification, feasibility and usefulness of persuading reluctant respondents" are able to persuade respondents at higher rates (Sinibaldi, et al., 2009). This paper will take a deeper look at the findings from the NSDUH surveys and draw comparisons to those derived from the NatCen survey. Specifically we will examine field interviewer attitudes, response rates, and refusals rates by interviewer tenure, gender, region, and bilingual ability. As response rates decrease across household surveys and resulting costs increase, information gained through this analysis may assist in attracting and retaining interviewers who may be successful obtaining cooperation as a result of their attitude. This analysis may also lead to improved training in these techniques. NSDUH, sponsored by the Substance Abuse and Mental Health Services Administration, provides national, state and substate data on substance use and mental health in the civilian, noninstitutionalized population ages 12 and older. Approximately 67,500 NSDUH interviews are completed annually and 600 field interviewers are staffed on the study.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Smarter Surveys for Smartphones: Optimizing Data Collection Using Mobile Devices

What's the Score? Creating an Index of a Survey's Mobile-Friendliness

Nicole R Buttermore *GfK*

Jason Knight *GfK*

Frances M Barlas *GfK*

Randall K Thomas *GfK*

The ever-increasing proportion of respondents who complete online surveys using a mobile device presents a major challenge to survey researchers. Items designed for traditional online surveys completed via desktop or laptop do not necessarily render well on smaller mobile devices – especially smartphone screens. Some examples of items that create problems for respondents on smartphones are grids, constant sum, and conjoint tasks. We created a metric that assesses the mobile-friendliness of a survey on each respondent's individual device. The metric includes measures such as the amount of time for pages of the survey to load, the ratio of the height and width of the webpage content to the height and width of the user's screen, the number of elements that do not appear on a respondent's screen due to screen size (e.g., the number of grid items or radio buttons that run off of a smartphone screen and require the respondent to scroll in order to view), and the number of large graphics included in a survey. In addition, the presence of survey tasks such as open-ended responses, numeric entry of responses, and conjoint, or ranking tasks can be entered to determine total respondent load. We examined the predictive power of these metrics across a large number of studies ranging in sample size from 400 to 50,000 cases. We used regression analyses to determine the validity of this index by assessing survey duration, break-offs, and completion rates by device type. We found that task difficulty, number of screens, screen size, and load speed were all significant predictors of increased survey duration and were associated with higher rates of survey break-offs, especially among smartphone respondents. We suggest a hierarchy of factors that should be addressed to redesign surveys in order to make them smartphone friendly.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Survey Nonresponse in 2016 Election Polling

Did White Non-Response Cause Pre-Election Polls to be Wrong in the 2016 US Presidential Election?

Joshua D. Clinton *Vanderbilt University*

John Lapinski *University of Pennsylvania*

Were pre-election polls wrong because of non-response by Trump supporters? With response rates at historic lows, pollsters always face the potential for errors if those that choose to respond to polls differ from those that do not after controlling for observable features. Responses are weighted to reflect the population of interest, but this addresses the bias caused by non-response only if non-response is correlated with the demographics being used to adjust the sample. Given a political climate in which an increasing number of voters question the legitimacy of traditionally respected institutions such as political parties, the media, and pollsters we question the extent to which non-response may have contributed to the errors made by state-level polls in the 2016 presidential election. Using individual level data from pre-election polls and exit polls, we provide evidence that systemic non-response among some white voters appears to have contributed to polling errors. Even after removing the potential impact of likely voter models, the results of pre-election polls differ from those of exit polls in ways that are more than can be attributed to last minute changes in opinion. The fact that people answering pre-election surveys appear to differ in their opinions than those answering the exit polls raises important questions for what might be done going forward to avoid such problems in the future.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

The Climate Change "Conspiracy"

Comparative Knowledge and Attitudes Toward Climate Change

Meaghan McKasy *University of Utah*

Jessica Pechmann *University of Utah*

Julia Howe *University of Utah*

Vanessa Bailey *University of Utah*

The scientific community is in agreement that anthropocentric climate change is a reality (Ho, Detenber, Rosenthal, & Lee, 2014; Oreskes, 2004). Yet many Americans today still do not perceive climate change as a personal and immediate risk (Leiserowitz, 2005) or even a legitimate claim. A nationally representative survey, conducted in October 2015 by the Yale Center for Climate Change Communication, endeavored to analyze how the nation felt about issues surrounding climate change. This proposed research utilizes the national survey data as a baseline for comparatively gauging perceptions of climate change at a large university. The study was distributed during a two week period of time in April 2016 to students enrolled at the university through the online research company Qualtrics. The final sample size was 836. Preliminary results show that students were significantly more likely to believe in climate change, and subsequently human caused climate change, than the national sample. These findings are interesting to note as the university has a higher average student age and is considered to be located in a traditionally conservative and religious state. This allows for an interesting case study, which is uniquely different from other university campuses, to be compared to the nationally representative data. Additionally, the data appears to show a knowledge gap for students who believe in climate change yet are uncertain about how to take action. Finally, the data emphasizes the importance of academic mentors and journals to students as trustworthy disseminators of information. This research explores the potential reasons for the dramatic difference in university and national findings, as well as contributes analysis to ongoing climate change studies. The researchers will further analyze the results and discuss implications of the findings.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Survey Nonresponse in 2016 Election Polling

Sore losers: Determinants of participation in post-election surveys

Bradley Jones *Pew Research Center - Washington, DC*

Adam Hughes *Pew Research Center*

The source of instability in pre-election polls is subject to ongoing debate. Recent research (Gelman et al 2016) suggests that shifts in candidate support that political observers attribute to opinion change are actually an artifact of changing sample composition. But existing research does not isolate the mechanism that might lead partisans to “opt out” of surveys, nor is it clear whether changes in sample composition correspond with observable political events. To clarify this debate, we examine survey participation and sample composition in the wake of elections. Existing research shows that perceptions of electoral legitimacy decrease in response to unfavorable electoral outcomes (Anderson et al 2005). Similarly, we expect that partisans on the losing side in elections will be less likely to participate in post-election surveys. Using a range of pre- and post-election surveys from 1980-2012, we show that partisans whose favored candidate loses the election are less likely to respond to survey invitations. This effect is consistent over time and is large: we estimate that partisans on the losing side of an election are 3-5 percentage points more likely to refuse re-interview than those on the winning side. We also report the results of an experiment designed to mitigate this observed drop-off by those who support the losing candidate.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Internet and Mobile Data Collection: Compliance and Patterns of Use

Effect of augmenting end-of-day recall with passively collected contextual data from smartphones

Mashfiqui Rabbi *University of Michigan, Department of Statistics*

H. Yanna Yan *University of Michigan*

Predrag Klasnja *University of Michigan*

Susan Murphy *University of Michigan*

Nowadays smartphones are the most personal device we carry. Smartphones are also potent data collection tools, and researchers can use smartphones to easily reach respondents and collect self-reports in a timely manner and in real life situations (e.g., with ecological momentary assessments (EMA)). However, EMAs are intrusive and hard to maintain for longitudinal studies. An less-intrusive alternative is to recall different daily moments at the end of the day, but recalls can be error-prone. However, recalling a daily moment can be augmented with different contextual data, e.g., location, type of user activity and the background ambience, collected during the same moment. While such context data can be acquired passively without any human involvement, their inclusion during recall may assist better memory recollection and produce more precise measures of the specific moments, e.g., participants' whereabouts, what they are doing, who they are with and their emotional states etc. To this end, we created a mobile app called ReVibe, which assists the recall of different pre-selected daily moments with diverse contextual information such as location, activity level (e.g., sitting, walking or in transport) and ambient noise level (e.g., noisy, people talking nearby etc) co-occurring near the moments. We hypothesize that recalls with such contextual information can retrieve momentary experiences more effectively than traditional recalls without the contextual information. We explore the feasibility of our approach in a 2-week pilot study (N=55). We ask questions about emotion, stress level, mindfulness and their whereabouts during the recall. We validate the quality of the recall by diversity and specificity of the responses, as well as by comparing the recall responses with in-the-moment EMAs. Results from this work will demonstrate a novel data collection approach about daily moments that requires low burden while reducing recall bias.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Apt. 3, We Pick You. Resident 1, We Pick You Too: Improving ABS Surveys and Household Selection Methods

Propensity Stratification with Auxiliary Data for Address-Based Sampling Frames

Jamie Ridenhour *RTI International*

Joseph McMichael *RTI International*

Auxiliary information appended to Address-Based Sampling (ABS) frames can be incorporated at the sample design stage to increase the likelihood of sampling addresses with target groups of interest; however, this information needs to be accurate and nonmissing. For much of the auxiliary data currently available the drawbacks are that its accuracy is unknown and it is missing for many addresses on the sample frame making it difficult to use for stratification. In this paper we discuss how we solve this problem by using data collected from a nationally representative household survey of youths ages 11-16 to create a propensity model for stratification. This propensity model is then applied to the address frame for a subsequent survey targeting the same age group in rural areas and stratifies those addresses by how likely they were to have members of the eligible population. We will discuss the impact this approach had on survey efficiency, both in terms of field cost and variance.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Smarter Surveys for Smartphones: Optimizing Data Collection Using Mobile Devices

Opportunities and Challenges: Using Smartphones and Mobile Devices for Innovative Data Collection

Carli Lessof *National Centre for Research Methods*

Patrick Sturgis *Univ. of Southampton*

There is a growing body of evidence about the completion of routine web surveys on smartphones, providing survey practitioners with specific aims such as raising response rates which are low relative to web survey completion on larger and fixed devices, reducing break off rates and ameliorating the effect of longer completion times. Furthermore, the gaps in evidence point survey methodologists to increasingly specific questions which could be addressed through further experimentation. There is much less clarity, however, about mobile surveys which take advantage of the additional functionality offered by smartphones such as cameras, sensors, Bluetooth, GPS and the more complex, integrated designs that are made possible by using Apps. The number of examples is increasing but remains small and those that have been reported are diverse in terms of design and purpose. This means that the promise of these more innovative approaches – that they will answer unresolved and emerging scientific questions because they operate closer to the moment of experience, or at times and places not previously possible, or collect new types of data – is still largely untested. This paper will draw on the existing literature and more recent examples to consider the available evidence and will draw out key themes and challenges such as (1) ethical issues and consenting processes, (2) respondent burden and the potential for these to be offset by a more engaging survey experience, (3) initial participation rates and whether these are sustained over the extended time frames often involved and (4) data quality, including whether new data types map successfully onto the concepts they are intended to capture and whether these are accurately collected. Finally, we consider some of the challenges inherent in this kind of survey innovation and a potential research agenda for new measurement.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Survey Nonresponse in 2016 Election Polling

Why nobody saw Trump coming: Non-response bias among non-college educated whites

David Shor *Civis Analytics*

Charlotte Swasey *Civis Analytics*

In contrast to nearly all pre-election polls and forecasts, Donald Trump won the 2016 presidential election. We utilize a novel dataset of over 100,000 polling responses and millions of polling attempts taken over the 2016 cycle to show that the bulk of the polling error this cycle can be explained by a strong positive correlation between propensity to answer phone surveys and support for Donald Trump among non-college educated whites. We also utilize historical data to show that this correlation is new to this cycle and presents a unique challenge to measuring partisan attitudes moving forward.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Hello! I am Your Interviewer and I May or May Not Affect Your Data Quality

Assessing the Reliability of the Massey-Martin Scale on ANES

Vanessa Meldener-Harrell *Westat*

Michelle Amsbary *Westat*

Brad Edwards *Westat*

Over the last decade, the Massey-Martin scale has been used in several surveys as a one-time measurement of respondents' skin color by interviewers. For the first time, the scale was used in the 2012-2014 panel for the General Social Survey (GSS), allowing the assessment of the reliability of the measurement over time and across interviewers. The analysis was conducted by Hannon and DeFina and suggests low intercoder reliability among interviewers in rating respondents' skin color. The American National Election Studies (ANES) has incorporated the Massey-Martin scale in its Time Series Study protocol. The Time Series Study is comprised of two components, a pre-election interview administered within the weeks prior to Election Day and a post-election interview administered within the weeks after Election Day. For the first time in 2016, the Massey-Martin scale was used in both components. Two main changes in the administration procedures existed between how the scale was used in GSS and how it is used in ANES. First, while the interviewer does not provide the rating in the presence of the respondent, ANES displayed the image of the scale on a CAPI screen, as a reminder of the rating levels. This differs from other administration procedures that suggest the interviewer memorizes the scale and references it only as needed. In addition, ANES had the interviewers rate their own skin color before the beginning of data collection, with the expectation that interviewers would use it as a reference point to assess the respondents' skin color. We analyze the skin color ratings between the pre- and post-interviews within the same interviewers and between different interviewers. We compare the results and assess the reliability of the scale and its administration for ANES.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Hello! I am Your Interviewer and I May or May Not Affect Your Data Quality
How do Low Versus High Response Scale Ranges Impact the Administration and Answering of Behavioral Frequency Questions in Telephone Surveys?

Mazen Sarwar *University of Nebraska-Lincoln*

Kristen Olson *University of Nebraska - Lincoln*

Jolene Smyth *Univ. Nebraska - Lincoln*

A consistent questionnaire design finding in self- and interviewer-administered modes is that low versus high response scale category ranges in behavioral frequency questions affect response distributions (Schwarz et al. 1985; Rockwood et al. 1997; Smyth et al. 2007). However, we don't know if response scale ranges affect responses when the interviewer is not instructed to read the response options to the respondent. In this case, if interviewers administer the question as instructed, the scale should not impact responses; however, interviewers may deviate from instructions about reading response options, probe using specific scale points, or otherwise adjust their behaviors in ways that communicate the scale to respondents and thus potentially impact responses. This paper uses data from an experiment on four behavioral frequency questions in which high and low response scale ranges that were not to be read to respondents were compared to examine (1) whether the ways interviewers administer questions differ across the scale ranges, (2) whether response distributions differ across the scale ranges, and (3) whether any effects in administration or distributions are moderated by interviewer experience or respondent age or education. The data are from the Work and Leisure Today II telephone survey (conducted summer of 2015, n=911, AAPOR RR3=7.8%). Preliminary analysis shows few differences in response distributions across the low and high frequency scales but the effect of the response scale on the number of conversational turns used to administer and answer the question is moderated by respondent education. Further analyses will examine differences across treatments in whether the interviewer read the question with or without changes, used probes, provided clarifications, verified answers, and recorded answers accurately. Response time will also be compared across versions. The paper will discuss implications of findings for questionnaire design and interviewer training.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Hello! I am Your Interviewer and I May or May Not Affect Your Data Quality
Survey-specific vs. Multi-survey Interviewers: How Does Working Concurrently on Multiple Surveys Affect Data Quality?

Holly Fee *U.S. Census Bureau*

Matthew Marlay *U.S. Census Bureau*

Jason Fields *U.S. Census Bureau - Washington, DC*

Nonresponse on a Computer-Assisted Personal Interviewing (CAPI) survey may undertake two forms. One form, unit nonresponse, is complete nonresponse to a survey when either a respondent does not agree to participate or an interviewer is unsuccessful when trying to locate a sampling unit. The second form, item nonresponse, occurs when a respondent completes the survey but he or she neglects to complete all applicable survey items. Unit and item nonresponse may induce nonresponse bias in survey estimates. Altogether, unit and item nonresponse rates are two key determinants of data quality. Factors that affect unit and item nonresponse have been linked to both respondents and interviewers. Research has shown that response rates differ by respondent attributes, such as age, gender, and income. Respondents are also more likely to have higher item nonresponse rates if he or she perceives certain survey items as sensitive. Observable characteristics of the interviewer, such as race and gender, are also associated with response rates. In addition, an interviewer's experience affects his or her skill and ability to gain cooperation from a sampling unit. Further, experienced interviewers may utilize more probes to help reduce item nonresponse. An interviewer characteristic that is less well understood is interviewer burden. One way to measure interviewer burden is the number of cases across surveys. This is particularly salient for those who work for the U.S. Census Bureau because many interviewers work multiple surveys. This paper examines how data quality of Survey of Income and Program Participation (SIPP) differs between interviewers who work just SIPP and interviewers who work concurrently on other surveys using data from Wave 1 of the 2014 SIPP panel.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Apt. 3, We Pick You. Resident 1, We Pick You Too: Improving ABS Surveys and Household Selection Methods

Improving Demographic Information for Address Based Sampling (ABS) frames

Joseph McMichael *RTI International*

Jamie Ridenhour *RTI International*

Address Based Sampling (ABS) frames are well suited for auxiliary data appends. Either through geocoding or direct address matching survey researchers are able to append ABS frames with everything from publicly available census data to more proprietary data sources compiled and sold by data clearinghouses like Experian, Acxiom, and KBM Group. If sufficiently accurate and complete these data sources can improve the efficiency of our sample designs, reducing cost and/or variance. However, the key word here is sufficiently accurate and the problem we often face is public data is too coarse and proprietary data is incomplete or inaccurate. This paper discusses how we employed machine learning algorithms to improve the demographic information on RTI's ABS frame. Using outcomes collected from a mail survey with more than 18,000 respondents we are able to model characteristics like adult age category, child age category, income and education. We then test the improved demographic profiles against two surveys to demonstrate its potential.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

The Climate Change "Conspiracy"

Misperceptions of Public Opinion: Americans Underestimate Belief in Global Warming

Adina Abeles *Stanford University*

Lauren Howe *Stanford University*

Jon Krosnick *Stanford University*

Bo MacInnis *Stanford University*

Perceptions of public opinion do not always reflect the actual opinions of the public. We document public misperception of public opinion on global warming. In a cross-sectional survey of a national probability sample of American adults in 2012 (N=804), respondents were asked to state whether they thought global warming has been happening, and to estimate the percentage of Americans who believe that global warming has been happening. Respondents underestimated levels of belief in global warming among all Americans (by 20%), Democrats (by 22 %) and Republicans (by 12.4%). This underestimation also surfaced when examining political partisans' perceptions separately. Both Democrats and Republicans underestimated public belief in global warming by similar amounts. Building on past research, we then explored possible sources of this misperception. First, we found that people who believed global warming was happening estimated higher levels of public belief in global warming when compared to people who did not believe global warming was happening. Second, consistent with recent studies demonstrating that Americans see the American public as more divided on issues than in actuality, we found that respondents estimated that the country is divided on the issue of global warming, perceiving only 56% of the country to think that global warming was happening (as opposed to the 73% that actually thought that way). Additional exploratory analyses can shed light on how people form perceptions of public opinion. We hypothesize that strong partisans will be more likely to overestimate the prevalence of their own belief among co-partisans, while simultaneously underestimating the prevalence of one's own belief among those in the other party. Misperceptions of public opinion can impact behavior and policy decisions, underscoring the importance of understanding how the public forms its perceptions of public opinion.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Hello! I am Your Interviewer and I May or May Not Affect Your Data Quality
The Socially Desirable Voter? A Multi-Mode Exploration of Race and Gender of Interviewer Effects in the 2016 Presidential Election

Ashley Koning *Rutgers University*

Debra Borie-Holtz *Rutgers University*

Kathleen Rogers *Rutgers University*

GraceAnn McMillan *Rutgers University*

Seminal texts in public opinion research have addressed such issues as “the American voter,” “the changing American voter,” the “unchanging American voter,” and “the vanishing voter.” This most recent election cycle brought to light another kind of voter – one whose self-reported electoral choices may have been influenced more by a heightened climate of race and gender issues instead of a true depiction of how one actually felt. This type of voter may have thus been reluctant to express certain attitudes or may have chosen to answer questions in ways deemed socially acceptable ... all because of the voice at the other end of the phone. Past interviewer effects research has addressed respondent interaction with interviewers of certain races and of a certain gender, and the interviewer effects research specifically on candidate choice has focused on the social and racial identity of the candidates in question themselves – such as the Wilder Effect, the Bradley Effect, and the Whitman Effect. But the unprecedented nature of the 2016 election cycle had unprecedented consequences for interviewer-responder interaction, especially regarding President-elect Donald Trump. Throughout the course of the general election campaign, we explored gender and race of interviewer effects on the 2016 race using a diverse pool of telephone survey interviewers, mainly drawn from the majority-minority Rutgers University student population. We assessed reported 2016 attitudes by interviewer gender, as well as by both the interviewer’s actual race and the interviewer’s race as perceived by the respondent. We then compared reported attitudes in live telephone interviews to those done through text and email at a similar time. This research has important implications for interviewer effects and interviewer populations, as well as social desirability bias.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Stemming the Tide of Nonresponse: Examining Study Design Characteristics that Influence Response Rates

Judging a Survey by its Envelope: Differing Results from Qualitative and Quantitative Research

Kay Ricci *Nielsen*

Lauren Walton *Nielsen*

Robin Gentry *The Nielsen Company*

Outer packaging of mail surveys is a researcher's chance to engage the potential respondent with images, color, size, or messaging. Private, government, and academic organizations all send out mail surveys and without a successful outer package, they may be doomed to the recycle bin. This research begins with a qualitative assessment of survey packaging using both a mail-sorting task and rankings with 30 in-depth interviews. A variety of materials were sorted along with "normal" mail to collect feedback on people's willingness to interact. We evaluated larger premium envelopes, similar in size to priority mail envelopes, with differing color and messaging (3 versions). Five 9X6 envelopes were included with differing color (manila vs. blue), more marketing feel vs. more academic, and thicker card stock vs. normal weight paper. Additionally, we explored willingness to interact with smaller letter envelopes (5 versions) and postcards (5 versions). When ranking, participants were asked to lay out each category of material (e.g. 5 letter size) that they would be most likely to least likely to open or notice. We asked participants to comment on the color, size, and texture of the envelopes, and choose the one they preferred. This research will share findings on the multiple aspects of envelope design to inform future design decisions. This qualitative research indicated a blue 9X6 inch envelope was preferred (10/15 participants). Furthermore, six participants ranked it as the best envelope compared to other four 9x6-envelope designs. Armed with these findings we tested the small blue 9X6 envelope against the large blue priority mail sized cardboard envelope in the summer 2015 Nielsen TV diary survey. Results indicate statistically lower survey returns with the smaller blue paper envelope compared to the larger blue cardboard envelope (16.8% vs. 19.2%). These results illuminate disconnect in qualitative and quantitative findings.

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

How Good Is Cheap? Evaluating the Quality of Estimates Derived from Non-Probability Samples

Internal validity and online panels: Comparing the social determinants of health

Nicholas Biddle *Australian National University*

Jillian Sheppard *Australian National University*

The range and diversity of options for conducting social research online has increased significantly over the last decade, including within Australia. Coinciding with increasing costs of traditional face-to-face or telephone surveys, a range of online panels are now available to draw sample from and conduct research on. Internationally, a large amount of research has been undertaken to test whether summary statistics from such panels exhibit particular biases relative to other data collection methodologies, and whether there are differences depending on the way the sample is selected (including probabilistically versus non-probabilistically). However, there has been less research investigating whether the relationship between variables exhibit similar biases. For much of the applied social research, the main question of interest is not whether a particular outcome has a given value or not, but what the relationship is between that outcome and a key policy parameter whilst holding other characteristics constant. In this paper, we use data from the 'Online Panels Benchmarking Study' to test whether the social and demographic determinants of key health outcomes vary across panel populations relative to other high-quality collection methodologies

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Internet and Mobile Data Collection: Compliance and Patterns of Use

Trends in User-Created Online Surveys

Jillesa Gebhardt *SurveyMonkey*

Laura Wronski *SurveyMonkey*

Online survey platforms have made it possible for anyone to quickly and cheaply create, send, and analyze data from their own surveys. Every day, SurveyMonkey users around the world send out 24,000 surveys to their friends, family, coworkers, clients, and customers. While some survey creators are aware of best practices and trends in survey methodology, most survey creators are not experts, and are using their best judgement to decide how to structure their survey. Meanwhile, 3 million respondents per day take a survey created on SurveyMonkey. How has the mode of taking surveys - mobile vs desktop, Android vs iOS - changed as the world has changed? Have non-expert survey creators picked up on these changes in the world, and have they adapted the surveys they create to accommodate? This research looks at the characteristics of non-expert created surveys over time and how the experience of survey takers may have changed as well. We will examine current trends in surveys, from the perspective of both the survey creator and survey taker. Using our database of surveys completed over the past three years, we will produce estimates of average survey length, average completion time, and survey topic by month. This will demonstrate whether surveys are currently longer or shorter, whether they take more or less time, or whether they are on a more diverse set of topics than they have been in the recent past. We will look at the growth of online survey-taking on mobile devices and by device in general. Finally, we will compare our results for surveys taken in the US with surveys taken around the world, to produce a true profile

May 21st, 2017
10:15 AM - 11:45 AM
Concurrent Session K

Internet and Mobile Data Collection: Compliance and Patterns of Use
Obedience in Respondents: Characteristics of Compliance

Yifei Liu *GfK*
Frances Barlas *GfK*
Nicole Buttermore *GfK*
Randall Thomas *GfK*

In survey design, there are a number of situations in which we want respondents to comply in order to obtain more valid data. One such situation occurs when we want people on their mobile device to view the survey in landscape mode rather than portrait mode due to the orientation of the questions and responses. Another case arises when the survey design will not work well on a smartphone and we request that respondents take the survey on a desktop/laptop. Finally, there are many cases where we want respondents to read over information concerning the study in order to obtain informed consent. We were interested in the extent to which respondents comply with survey requests as well as the characteristics of respondents who are more or less likely to comply with such requests. We looked at data from three separate studies – the first asked for screen compliance, the second asked for device compliance, and the third asked for informed consent. For the first two, we examined whether respondents complied based on our survey software detection for screen orientation and survey completion device. For assessing compliance with reading informed consent, we examined speed of reading based on the time to complete the response, with the fastest responses indicating a much lower likelihood of having read the consent section in its entirety. Across all three studies, we found that respondents who were younger (less than 29) or males were less likely to comply. Education had some paradoxical effects, with lower education (High School or less) and higher education (Graduate school or higher) having the lowest compliance on the consent task. We review some alternatives to improving compliance for these harder to cooperate groups.