



**AAPOR**

AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH

**69th Annual  
Conference**



**Measurement  
and the Role of  
Public Opinion  
in a Democracy**

**2014 Conference Abstracts**

**May 15 – 18, 2014**

Anaheim Marriott • Anaheim, CA

**Thursday, May 15**  
**1:30 p.m. – 3:00 p.m.**  
**Concurrent Session A**

## **Moving Surveys to the Web**

### **The Effects of Adding an Internet Response Option to the American Community Survey**

Debbie Griffin, *U.S. Census Bureau*

Stephanie Baumgardner, *U.S. Census Bureau*

David Raglin, *U.S. Census Bureau*

The American Community Survey (ACS) is a large household survey that continuously collects data on social, demographic, economic, and housing characteristics. Traditionally, the ACS collected data using three sequential modes—a mailout-mailback paper questionnaire, a telephone followup and a personal visit followup. In 2013, the ACS added a fourth mode—collection through an Internet instrument. Sample addresses now receive an initial invitation to complete the survey via Internet, with nonrespondents eligible for the other three modes. This paper describes the full implementation strategy used in the 2013 ACS and summarizes survey response by mode, and how that differs by area. The paper also compares the combined Internet and mail self-response rates with the mail response rates in 2012 and the rates from the Bureau's 2011 Internet test. Analysis of the characteristics of Internet respondents is also presented.

### **Results from Testing a Web Mode for the Consumer Expenditure Diary Survey**

Ian J. Elkin, *Bureau of Labor Statistics*

Laura Erhard, *Bureau of Labor Statistics*

Brett McBride, *Bureau of Labor Statistics*

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

The Consumer Expenditure (CE) Diary Survey relies on a pencil-and-paper instrument to collect data from multiple household members. There are a number of drawbacks to this method, such as limiting entry to a single diarist in a one location, who must carry the diary with them throughout the day. A web diary has the potential to address these limitations and, over time, the feasibility and potential benefits of using a web diary have grown. There is evidence to suggest that accuracy of reporting may be improved through the use of a web diary (Couper, 2008). A web mode also has the potential to improve unit and item response rates, and would allow for easier access across consumer unit (CU) members since respondents can enter expenditure data from any internet-enabled location and would not be tied to a single instrument kept in one location. Finally, a web component has potential cost savings over PAPI due to reduced or eliminated materials, scanning, and data entry expenses. Beginning in January 2013, CE fielded a test to determine the feasibility of collecting expenditure data via a web diary. In this presentation, we report on findings regarding sample performance, sample composition, diary maintenance, and other associated operational issues as well as provide recommendations for implementation and further research.

## **Challenges and Strategies Involved in Adapting a Very Large-Scale Survey for Online Administration**

Rossi Dobrikova, *Experian Marketing Services*

Christine Kudish, *Experian Marketing Services*

Susan Sanford, *Experian Marketing Services*

Max Kilger, *Experian Marketing Services*

While a lot of research has been focused on strategies for adapting traditional paper and pencil surveys to online, much less work has been conducted on the adaption of very large (120+ page) surveys for online administration. This paper addresses the methodological challenges and associated strategies being deployed for adapting a very large-scale survey for online administration, in an effort to provide multi-mode opportunities for lengthy paper surveys to respondents. It explains our approach to questionnaire and survey navigation design and addresses the challenges of reducing respondent burden, while exploiting some of the benefits of online administration for improving data quality such as item nonresponse rates.

A number of key issues are discussed, including strategies for splitting the survey into modules, the associated design of navigation among those modules, the issues involved in providing progress feedback to the respondent, and the challenges revolving around devising the most effective means for encouraging respondents to return to the survey after temporarily leaving the survey. Additionally, we discuss the design and implementation of “soft controls” for critical gate questions, where, if the respondent moves on without answering the gate question, a soft control intervention is deployed to encourage the participant to respond to the gate question. Finally, we discuss some of the design challenges for very large surveys, where traditional wisdom about online survey administration such as scrolling may not always apply.

This paper features some of the critical design decisions involved in the monumental task of adapting a very large survey to the online environment, and is meant to be the first in a series of papers as we test these design strategies and report on the results of the different approaches.

## **In-Person or On-line? The Future of the American National Election Study**

Brian F. Schaffner, *University of Massachusetts, Amherst*

Stephen Ansolabehere, *Harvard University*

The American National Election Study (ANES) has been the “gold standard” of public opinion surveys since its inception in the 1940s. The survey has traditionally been conducted utilizing face-to-face interviews, but the rising cost of this method coupled with technological advancements in online survey methodologies have provided an impetus for considering how the survey will be conducted in future years. In 2012, the ANES content was replicated in two separate on-line surveys—one utilizing Knowledge Networks/GFK and the other utilizing You Gov, America—in addition to the traditional face-to-face format. This paper provides a detailed comparison of how the two on-line surveys fared compared to the face-to-face content. In doing so, it provides guidance for scholars evaluating how the ANES should be conducted in the future.

## **Surveying Low Income Parents: To Web or Not to Web, is that the Mode?**

Sara Skidmore, *Mathematica Policy Research*

Cassandra Meagher, *Mathematica Policy Research*

Jerry West, *Mathematica Policy Research*

Administering brief, reliable measures of children's development, behavioral health, and family context in a cost efficient manner is important for conducting studies of low-income children and families. Mathematica Policy Research, with funding from the Administration for Children and Families, is conducting a pilot study to test the feasibility of using a web option with a low-income population. The study will administer the Survey of Well-Being of Young Children (SWYC) to 450 English- and Spanish-speaking parents of Head Start children to compare web and phone administration of the instrument. The SWYC is designed to be completed by parents in ten minutes, and to measure three domains of children's functioning: (1) developmental, (2) social/emotional, and (3) family context. Parents will be randomly assigned to complete the survey on the Web or via Computer-Assisted Telephone Interview (CATI). The study will look at response rates, timing and completeness of responses by mode, and will include cognitive interviews conducted with respondents of both modes to identify challenges parents experienced in completing the survey, and to assess the feasibility of using the Web in large-scale studies of low-income families with young children. The analysis will also examine the range of responses and scores, internal consistency, and correlations between individual items and total scores that provide information about children's development and functioning across the full sample, and separately by mode.

## **Social Issues, Public Resilience and Public Resistance**

### **Measuring Recovery and Resilience after a Disaster: A Survey of the Hurricane Sandy Region 6-Months after the Storm**

Trevor Thompson, *NORC at the University of Chicago*

Jennifer Benz, *NORC at the University of Chicago*

Becky Reimer, *NORC at the University of Chicago*

Emily Alvarez, *NORC at the University of Chicago*

Daniel Malato, *NORC at the University of Chicago*

Striking landfall in the U.S. on October 29, 2012, Superstorm Sandy affected large areas of coastal New York and New Jersey, devastated communities, killed more than 130 people, and caused tens of billions of dollars in property damage. Anecdotal evidence and reporting suggested the impacts of the storm were still being felt in many communities months later. The high level of damage and often slow recovery from natural disasters such as Superstorm Sandy are causing policymakers and emergency managers to ask: what characteristics of a community make it more resilient, and thus better able to effectively respond? The AP-NORC Center, with funding from the Rockefeller Foundation, conducted a wide-ranging research project on community resilience, addressing the question of why some individuals and communities recover more quickly and more effectively in the face of a natural disaster. With an objective of providing systematic data to address this question, we conducted a national survey of 2,025 individuals, with an oversample of 1,007 interviews with residents of the Superstorm Sandy affected region. By collecting rigorous data on neighborhood characteristics and social factors, the study provides an opportunity to understand how these factors relate to recovery, resilience, and perceptions of individual and neighborhood-level vulnerability to future disasters. Findings reveal that storm victims turned to family, friends and neighbors for support or assistance more

often than formal sources for support or assistance. Furthermore, neighborhoods lacking in social cohesion and trust more generally had a more difficult time recovering from Sandy. Individuals in slowly recovering neighborhoods were more likely to report greater levels of hoarding food and water, looting and stealing, and vandalism in their neighborhoods during or immediately after the storm.

### **Safety and Solidarity After the Boston Marathon Bombing: A Comparison of Three Diverse Boston Neighborhoods**

Jessica L. LeBlanc, *Center for Survey Research, University of Massachusetts Boston*

Philip S. Brenner, *Department of Sociology, University of Massachusetts Boston*

Anthony M. Roman, *Center for Survey Research, University of Massachusetts Boston*

Naa Oyo A. Kwate, *Dept. of Human Ecology and Africana Studies, Rutgers University*

On April 15, 2013, residents of Boston, Massachusetts witnessed the bombing of the annual Boston Marathon. The timing of this tragic event provided a unique opportunity to study the impact of a terrorist attack on city residents. This paper investigates the effect of the bombing on residents of Boston, focusing on the meso-level — how the tragic incident changed, or did not change, how Bostonians live in and feel about their community and neighborhoods — to address a relatively under-investigated level of analysis in the research literature. Unlike prior research that began weeks or months after a terrorist attack and used retrospective reports, this study spans the focal event. Data from an ongoing study investigating Boston residents' subjective valuations of their city and their neighborhoods, perceived inequalities, and quality of mental and physical health was used in these analyses. An address-based sample of residents from three neighborhoods, distinct in racial and economic makeup was surveyed by mail using a three-contact protocol. About two-thirds of respondents answered the survey before the bombing (n=581) and slightly over a third answered the survey after the bombing (n=313) for an overall response rate of 38% (AAPOR RR4). Assessments of safety, city and neighborhood satisfaction and cohesion, mental health, and other key measures vary greatly between the three neighborhoods, which are diverse in racial and economic composition, but also vary in proximity to the bomb site. Net of neighborhood differences, the bombing had a strong negative effect on neighborhood cohesion and reduced use of public transit. Strong interactions are also found between timing of survey completion (pre and post bombing) and neighborhood for feelings of neighborhood safety and cohesion and use of public transportation, but not mental health status.

### **Colorado: Gun Control Legislation and Recall Elections The Influence of Polls in Policy and Politics**

Floyd Ciruli, *Ciruli Associates*

Gun control politics in Colorado in 2013 provides a powerful backdrop to investigate the impact of public opinion and polling in the passage of historic gun control legislation and the subsequent backlash of voters in two successful legislative recall elections. Colorado's politics had significant state and national implications for gun legislation. As a swing presidential state, Colorado is carefully watched as to its opinion dynamics for clues to shifting national trends.

This presentation will analyze the polls, their media coverage and the statements of political actors concerning them for their impact on the legislative and political process.

Specifically, this proposal will present data and analyses of differential issue intensity, decay and partisan distribution. The recall election poll, in particular, reaffirmed the long observed

principles about the differences between a passionate opposition vs. more diffuse voter support. It also highlighted an issue that cut across partisan lines. And it reinforced the principle that an issue's intensity can decay rapidly from the horrific incidents that initially generate political action.

Polling and public opinion was a significant factor in each step of the 2013 gun control battle in Colorado. From polls published in early January that encouraged Democratic legislators and a moderate Democratic governor to move forward with gun registration and magazine limitations, to statewide polls in early and mid-summer showing a change in the state's political mood and downward shift in the governor's popularity, to a poll in one of the legislative recall districts showing an upset, a substantial amount of polling either motivated, warned or described the politics and public opinion of gun control in Colorado.

### **The Measurement and Influence of White Racial Sympathy in American Politics**

Jennifer Y. Chudy, *University of Michigan*

Since The Authoritarian Personality, social scientists have studied the role of out-group animus in social life. However, few researchers have considered the role that out-group sympathy might play in informing social behavior. In this paper, I examine racial sympathy among Whites by documenting its prevalence, antecedents, and political consequences. I start by offering a critique of previous measures of the concept, most of which have inverted measures designed to calibrate resentment. I then examine the antecedents of racial sympathy including motivation, empathy, social desirability, stereotype threat, and dimensions of personality (openness and agreeableness). I subsequently develop my own measure, which uses a series of vignettes depicting racial discrimination, to gauge respondents' sympathetic responses to discrimination. I demonstrate the validity of the measure through a series of tests. Using this measure, I find that a non-trivial portion of Whites carries racially sympathetic views and that these views influence their attitudes towards Black candidates and implicitly and explicitly racial public policies. I also demonstrate that this measure provides exploratory power beyond that which is offered by current measures of racial animus. This paper adds conceptual richness and an empirical demonstration of a powerful force in American politics, most often and inadequately defined in terms of its opposite.

## **Methodological Briefs: Survey Participation**

### **Do Hard-to-Interview Groups Self-Respond More When Offered an Internet Reporting Option?**

Rachel Horwitz, *U.S. Census Bureau*

Elizabeth Nichols, *U.S. Census Bureau*

Jennifer Tancreto, *U.S. Census Bureau*

With increases in Internet penetration across the country and rising survey costs, in 2013 the American Community Survey (ACS) added an Internet response mode. In the ACS design, respondents first receive a mailing that tells them how to access the Internet instrument to complete the survey online. Households that do not respond then receive a paper questionnaire in the mail several weeks after the invitation to complete the survey online. Households that do not respond by paper or Internet are then contacted by an interviewer via telephone and if they still do not respond, a subset is selected to receive a personal visit from an interviewer. Demographic and Internet access characteristics are known indicators of response patterns.

Specifically, households with children under age five, households in which all residents are under age 30, households with a Black respondent, a Hispanic respondent, or a respondent with less than a high school education, households in which at least one resident speaks Spanish, and renters are traditionally harder to interview than other demographic groups. While some of these groups have high Internet penetration, suggesting they may be easier to interview via Internet than the traditional ACS modes, other groups have low Internet penetration, which may make them even more difficult to interview. Therefore, this paper will assess the impact of offering an Internet option prior to a mail option on the self-response rates of these hard-to-interview groups. Additionally, we will measure the mode preference among these groups and determine whether this preference is associated with changes in Internet and paper response rates.

### **Would You Pick Up the Phone?: The Utility of Local Respondent Telephone Numbers on Caller ID Displays**

Alyson Croen, *NORC at the University of Chicago*

Elizabeth Shenkman, *University of Florida*

Jacquelyn George, *NORC at the University of Chicago*

Martin Barron, *NORC at the University of Chicago*

Kimberly Case, *University of Florida*

Unfamiliar caller ID area codes have the potential to suppress survey response by allowing potential respondents to screen out unwanted or unknown calls. This may be especially damaging when a survey organization is contacting individuals in other states as respondents may consider calls from unknown area codes as more likely to be unwanted solicitations.

In 2013, NORC at the University of Chicago conducted a caller ID experiment on a subset of the sample from the Wellness Incentive and Navigation (WIN) project. WIN is a three-year longitudinal project with an annual survey component, designed to improve the experiences of Medicaid beneficiaries in Texas. In our experiment, half of the respondents were sent a local area code during dialing for their respective service area while the other half of the respondents were shown a Chicago-area area code, which is the location of NORC's call center. This experiment was created after assessing the impact of switching to a local number midway through recruitment for the Main WIN survey, implemented the year prior in 2012.

Preliminary results show that the local number did not appear to improve response rates. This presentation will explore the data surrounding the performance of the experimental groups, delve into the impact of the local number during the initial implementation in 2012 and the current experiment, and will focus on considerations for future work since the WIN surveys are longitudinal.

### **So Nice They Respond Twice: What to do With Duplicate Responses**

Nikkilyn Morrison, *Mathematica Policy Research*

Nancy Duda, *Mathematica Policy Research*

Anna Situ, *Mathematica Policy Research*

Karen E. Bogen, *Mathematica Policy Research*

Researchers have to de-duplicate a data file when sample members return multiple completed questionnaires. When conducting a mail survey, the larger the sample, the greater the potential for duplicate responses because of the increasing lag between the time an address file of non-respondents is prepared and the follow-up mailing is assembled and shipped. During this lag

time, sample members may return a completed survey and then receive second mailing. Most respondents discard the second survey if they have already completed the first, but some respondents complete the survey again. For the Consumer Assessment of Healthcare Providers and Systems (CAHPS) surveys, the Agency for Healthcare Research & Quality (AHRQ) specifies that a researcher should accept the first returned survey and discard additional responses.

This paper uses data from a mail-only survey of over 153,000 adult patients conducted for the Evaluation of the Comprehensive Primary Care Initiative, sponsored by the Centers for Medicare & Medicaid Services, to assess the differences between duplicate responses to the same survey. The questionnaire contains the entirety of CAHPS Clinician & Group Patient-Centered Medical Home survey, among other items. Approximately 52 percent of sample members are Medicare beneficiaries. Most data collections do not have enough cases to assess the implications of the keep-the-first guideline. With our large sample size, we have over 600 respondents who have returned more than one survey, so we can test to see what difference it makes to follow the guideline. We investigate the effects of using alternative strategies and evaluate the completeness and order that surveys were returned to determine if the first or second is systematically more complete. We also review survey responses to see how duplicate responses differ on different types of survey items, such as sensitive personal information versus respondent opinions or nonsensitive factual information.

### **Providing a Deadline for Response: Results from Two Recent Experiments**

Ashley Kaiser, *American Institutes for Research*

Jill Walston, *American Institutes for Research*

Rebecca Medway, *American Institutes for Research*

Cong Ye, *American Institutes for Research*

Roger Tourangeau, *Westat*

Providing a deadline for sample members to respond could potentially have either a positive or negative effect on the speed with which survey responses are received and on the overall response rate. On one hand, letting sample members know that there is a short window of time for them to respond may lead them to respond quickly for fear of missing the opportunity to participate. On the other hand, providing a deadline may appeal to respondents' tendency to procrastinate right up until the deadline, thereby slowing the speed of response. A deadline also may suppress response rates if it leads respondents who feel pressed for time to discard the survey request because of the impression that they will not have a chance to respond before the imposed deadline. It also is reasonable that deadlines may affect sample members in different ways depending on characteristics such as topic salience. However, there has been relatively little research looking at the effect of deadlines on survey response – and most of that which does exist is limited to mail surveys.

This presentation will review the results of two recent experiments that varied whether survey sample members were given a deadline for responding. The first was a Web survey in which sample members were assigned to a deadline condition or no deadline condition. The second was a mixed-mode mail and Web survey in which sample members were assigned to (1) a short deadline condition, (2) a long deadline condition, or (3) a no deadline condition. We will report on the effect that the deadline had on the speed with which responses were received in these two experiments, as well as on the overall response rate. We also will discuss whether the deadline had an effect on sample composition and responses to key variables.



## **Underrepresentation of Young Women in Dual-Frame Surveys: Causes and Implications**

Eran Ben-Porath, *SSRS*

Susan Sherr, *SSRS*

Jordon Peugh, *SSRS*

For years, survey researchers have recognized that telephone samples tend to over-represent older, female respondents. The advent of cell phone interviewing has helped to overcome this bias, in that cell phone respondents are more likely to be young and male, thus resulting in a more representative sample overall. However, the expansion of cell phone numbers in dual-frame RDD samples, has introduced a different type of bias: underrepresentation of younger women. Whereas among older respondents, the share of women completing interviews is disproportionately higher than men's, the trend is reversed among younger respondents: both genders are under-represented relative to their share in the population, but women much more so than men. Using data from a weekly national omnibus survey and monthly media polls, we detail this trend, illustrate the impact of underrepresenting young women on outcomes such as party identification and presidential approval ratings, and stress the importance of correcting age and gender as a cross-product in sample weighting.

## **Obtaining Assent from Minors - Assessing Comprehension in the National Survey of Youth in Custody**

Tim Smith, *Westat*

Leanne Heaton, *Westat*

Sharon Zack, *Westat*

David Cantor, *Westat*

While most surveys are required to administer informed consent, there is very little data on how often respondents actually understand the survey conditions for participation. This is an especially important topic for surveys of vulnerable populations, such as minors. The National Survey of Youth in Custody collected data from approximately 11,000 youth about sexual contacts in juvenile facilities. This paper describes a procedure that both assessed and fostered youth comprehension of the assent. An interviewer read the assent to a youth and then assessed the level of comprehension based on the youth's responses to questions about the study procedures. If the youth failed to demonstrate understanding, the interviewer repeated and/or paraphrased the assent text related to the question and asked the question again. Youth that still failed to respond correctly were not asked to complete the survey. The findings show that simply reading an assent script to youth may not be sufficient in ensuring comprehension of key conditions related to research participation. Twenty-eight percent of youth did not initially understand at least one condition of the assent. After repetition and paraphrasing, less than 0.5 percent still did not understand all conditions. Youth sex, age, and race were significantly associated with comprehension. Those who needed assistance were less likely to complete the survey and more likely to have inconsistent responses within the survey. In response to debriefing questions about the survey experience, youth who needed assistance were also more likely to express difficulty understanding some survey questions. Analysis of interviewer characteristics showed that neither interviewer sex nor race was significantly associated with youth comprehension. However, interviewer ethnicity was predictive of Hispanic youth comprehension.

## **The Role of the Interviewer's Responsiveness in Avoiding and Converting Refusals**

Nora Cate Schaeffer, *University of Wisconsin - Madison*  
Dana Garbarski, *University of Wisconsin - Madison*  
Jennifer Dykema, *University of Wisconsin Survey Center*  
Douglas W. Maynard, *University of Wisconsin - Madison*  
Bo Hee Min, *University of Wisconsin - Madison*  
Ellen Dinsmore, *University of Wisconsin - Madison*

As Groves and Couper noted (1996: 67), success in recruiting sample members to participate in a survey interview may depend less on maximizing “the likelihood of obtaining a ‘yes’ answer in any given contact,” and more on minimizing “the likelihood of a ‘no’ answer over repeated turn-taking in the contact.” However, identifying specific behaviors of the interviewer that promote or avoid “no” is methodologically complex. When we observe a recruitment call, what we see results from both the interviewer’s actions and the respondent’s propensity to participate. For example, “maintaining interaction” may be associated with increased likelihood of participation because the respondent’s propensity to participate allows the interaction to continue. Schaeffer, Garbarski, Freese, and Maynard (2013) recently proposed examining a particular quality of rapport in the recruitment call, the responsiveness of the interviewer. One category of responsiveness consists of empathic utterances; such utterances may affirm the sample member’s stated position, display reflective listening, express empathy, acknowledge costs of participating, or maintain optimism about the desired outcome. Such actions may maintain interaction because responsiveness is linked to the motivation of the respondent (Cannell & Kahn 1968; Dijkstra & Van der Zouwen 1987) and social exchange (Dillman 2007). In contrast, when an interviewer uses “why not” and similar questions in an attempt to maintain interaction, the questions may lead to the “no” they seek to avoid. Our analysis examines these two paths through the recruitment call. We also analyze interviewers’ responsive behaviors in subsequent attempts to convert those who initially refused. We use audio recordings from the 2004 wave of the Wisconsin Longitudinal Study (WLS) and an innovative design that selects 257 paired declinations and acceptances matched on the sample members’ propensity to participate in the survey.

## **Influence of Prior Respondent-Interviewer Interaction on Disclosure in Audio-CASI**

Hanyu Sun, *Joint Program in Survey Methodology*  
Frederick G. Conrad, *University of Michigan*  
Frauke Kreuter, *University of Maryland*

Audio computer-assisted self-interviewing (audio-CASI) is widely used to collect information about sensitive topics such as illicit drug use or sexual behavior. Typically, respondents first answer questions about nonsensitive topics in a computer-assisted personal interviewing (CAPI) and are then switched to audio-CASI for sensitive questions. In audio-CASI, respondents read questions on a computer screen while hearing the same questions read aloud over headphones. The general finding is that audio-CASI increases disclosure of sensitive information relative to CAPI (e.g., Tourangeau and Smith, 1996). In these studies, audio-CASI is treated as an independent mode of data collection even though the audio-CASI module follows a CAPI module. None of the existing research has investigated the possibility that the interviewer-respondent interaction prior to the audio-CASI questions may affect disclosure in audio-CASI. The prior interviewer-respondent interaction may create a sufficient amount of social presence to reduce disclosure in audio-CASI. In particular, the respondent may build a

positive relationship or rapport with the interviewer during their prior interaction. Additionally, if the voice used in the audio-CASI audio-file sounds similar to the CAPI interviewer, it may remind respondents about the presence of the interviewer. Greater rapport as well as similar interviewers and ACASI voices may reduce disclosure in the audio-CASI module compared to lower rapport and more distinct voices. We test this in a laboratory experiment in which the respondent first completes a 30-minute CAPI interview – plenty of time to develop rapport—and then completes a 15-minute audio-CASI interview. We vary two factors: the level of rapport in the prior interaction (interviewers are either naturally high or low in rapport) and the similarity of the CAPI and audio-CASI interviewers' voices (same voice vs. different voice). We examine admissions of sensitive information across these four experimental conditions. The findings are interpreted in light of their implications for survey design.

## **Current Cross-Cultural Research on the Use of Paradata to Examine Multiple Error Sources Simultaneously in the Total Survey Error Framework**

### **Differential Response Styles of Subjective Life Expectancy and Cultural Differences in Time Orientation**

Sunghee Lee, *University of Michigan – Ann Arbor*

With an unprecedented population aging pattern, the subjective life expectancy (SLE) question has emerged as an important predictor of mortality especially for the older population. SLE asks respondents to estimate the probability that they will live up to a certain age. In spite of its predictive power for the general population, its applicability for racial/ethnic minorities in the U.S. has been questioned. This study hypothesizes that compared to non-Hispanics, Hispanics are likely to experience a higher level of difficulty in answering to SLE due to their time orientation manifested through item nonresponse and more likely to report a 0 percent probability due to their fatalistic view on life and health. Further, we hypothesize that these response patterns are more pronounced among Hispanics interviewed in Spanish than those interviewed in English by using interview language as a proxy measure of acculturation.

Using the Health and Retirement Study, we observationally examined these item response patterns of SLE and the relationships between SLE and actuarial life expectancy (ALE) based on the Life Table and between SLE and subsequent mortality status across racial/ethnic groups. We found 1) item nonresponse and 0 probability reports were highest among Hispanics mainly due to Hispanics interviewed in Spanish, 2) SLE item nonresponse was associated with a higher mortality rate, and 3) SLE was a poor correlate of ALE and a poor predictor of mortality for Hispanics.

### **A Longitudinal Analysis of Nonresponse and Linkage Non-Consent Bias in the German “WeLL” Study**

Joe Sakshaug, *Institute for Employment Research (IAB)*

Surveys are subject to multiple error sources which threaten the validity of inferences obtained from them. Much of the Total Survey Error literature has focused on identifying and measuring these error sources in cross-sectional surveys. Unfortunately, longitudinal surveys have not received the same attention. There are reasons why errors in longitudinal surveys may be more severe compared to their cross-sectional counterparts. For instance, in longitudinal surveys, the

original sample size typically decreases due to attrition, which may increase the effects of nonresponse bias over time. Other sources of error may also be problematic in a longitudinal context, including linkage non-consent. Many surveys ask respondents for permission to link their survey record to administrative databases, but not all respondents consent to this request. Not providing consent to record linkage has been shown to affect the validity of cross-sectional estimates based on administrative data, but its impact on longitudinal estimates has not been well studied. In this paper, we examine both nonresponse and linkage non-consent biases in the German Panel Study “Further Training as a Part of Lifelong Learning” (WeLL). WeLL is a telephone study of employees at certain establishments who are asked about their further training activities. Four annual waves of the study are considered here. Administrative data corresponding to these four waves are available and are used to assess the magnitude of nonresponse and linkage non-consent biases over time. We present bias estimates for several administrative variables and compare their individual magnitudes for each wave of the study. Practical considerations and some guidance for longitudinal surveys will also be provided.

### **Survey Response as a Process: How Response Propensity and Data Collection Costs Vary Together in a Face-to-Face Survey**

Andrew Mercer, *Westat*

Typically, response propensities are estimated as the probability that a respondent will cooperate at some unspecified point in the field period, while costs are frequently estimated in aggregate terms such as interviewer hours per completed interview. Though useful for planning purposes, this approach provides little help in guiding interviewer behavior in the field. This research treats survey response as a process rather than a single event. In face-to-face surveys, interviewers often have considerable discretion over when and how they attempt to make contact with sampled respondents. Different patterns of interviewer activity require different levels of time and effort. They also have different effects on response propensity. The expected cost effectiveness of any given action depends on history of work performed on a case as well as characteristics of the respondent. Using call history data from the Survey of Free Time (SOFT), we develop a multilevel discrete time hazard model of the response process that incorporates changes in respondent status that occur during the data collection process (such as when a respondent makes an appointment or initially refuses to participate). This model allows us to determine how different interviewer actions affect the probability of response given the history of interactions that occur between interviewer and respondent. Using predictions generated by these models, we explore the relationship between cumulative data collection costs and changes in response propensity over the course of data collection for different classes of respondent. We also investigate how different rules for determining when to stop work on a case can affect survey estimates and data collection costs.

### **Using Doorstep Concerns Data to Characterize and Correct for Nonresponse**

Ting Yan, *University of Michigan – Ann Arbor*

Shirley Tsai, *U.S. Bureau of Labor Statistics*

Doorstep concerns – one type of paradata – capture the interactions between interviewers and potential survey respondents during the survey introduction and reveal the concerns sampled members have expressed about the survey request and also their reasons for refusing the survey request when refusal occurs. We’ve created two parsimonious measures that retain the interrelationships inherent in the doorstep concerns data – the Perceived Concerns Index (through principal component analysis) and the Reluctance Class (via latent class analysis) – and have demonstrated that both measures are effective in characterizing and assessing the

level of reluctance exhibited by potential survey respondents. Taking advantage of the Consumer Expenditure Interview Survey, this paper continues the previous research by exploring the impact of reluctance, as measured by the Perceived Concerns Index and the Reluctant Class, on survey respondents' likelihood to participate in the later survey requests and the likelihood to provide missing data to survey questions. Furthermore, this paper will also explore the possibility of using both summary measures in nonresponse adjustment and in imputation. We will empirically examine the changes in estimates of key variables of interests after including measures of reluctance in nonresponse adjustment and imputation.

### **What Can Paradata Tell Us About Nonresponse Bias After Standard Post-Stratification Adjustments?**

Brady T. West, *Institute for Social Research*

Brian S. Kreuger, *University of Rhode Island*

The potential of nonresponse bias in survey estimates, engendered by decreasing response rates in surveys of all modes, is a ubiquitous challenge facing modern public opinion researchers. Paradata, or data describing the survey data collection process, show theoretical promise for addressing this widespread problem. Unfortunately, very few studies have systematically assessed the use of these data for post-survey nonresponse adjustments, and those studies that do exist generally do not identify paradata that correlate well with both response propensity and key survey variables. The advancement of original research on the utility of paradata for addressing nonresponse problems, especially in an era of "Big Data" where large volumes of paradata are being collected every day, relies on studies of the relationships of survey paradata with key variables and response propensity. By presenting analyses of data from the National Survey of Family Growth, this study uses a large set of paradata and auxiliary information to assess the potential of paradata as an independent tool for post-survey nonresponse adjustments. We begin by using this rich set of auxiliary information about respondents and non-respondents to obtain response propensity scores. We then show that the distributions of several respondent attitudes and behaviors vary significantly across quintiles of these propensity scores, after controlling for standard demographic factors that are often used in post-stratification adjustments. Finally, we show that standard post-stratification adjustments neglecting to account for these data and their relationships with response propensity can lead to substantively different estimates. Our results imply that standard post-stratification adjustments may not be entirely effective at removing nonresponse bias from all survey estimates, especially so for selected demographic subgroups, and offer compelling evidence that paradata and auxiliary data may be effectively used along with standard post-stratification adjustments to reduce nonresponse bias. We conclude with suggestions for practitioners and future research.

### **Maximizing the Quality of Reports: Design and Data Collection Strategies**

#### **A New Survey Tool for Ego-Centered Networks**

Tobias H. Stark, *Stanford University*

Jon A. Krosnick, *Stanford University*

Questions about survey respondents' social contacts (ego-centered networks) have become increasingly popular in surveys because research has shown that social contacts influence people's behavior and attitudes. How to collect information on a respondents' social network is

well understood when done during face-to-face interviews, and ego-centered network questions have been successfully implemented in, for instance, the General Social Survey or the American National Election Study. However, a number of studies have shown that data quality suffers if network items originally developed for face-to-face interviews are used in online surveys. Questions about ego-centered networks impose a high cognitive burden on respondents because the same questions must be answered for all network contacts. In online surveys, where no interviewer can motivate the respondents, ego-centered network questions lead to high break-off rates and more item-nonresponse, and respondents tend to nominate fewer network contacts.

This study presents a new data collection tool for ego-centered networks in online surveys. This tool reduces cognitive effort and increases respondents' motivation through two innovations. First, we make use of Web 2.0 tools to generate a more engaging experience for survey respondents when answering the same question for all of their network contacts. Respondents see a visual representation of their social network and answer questions by interacting with this network picture. The second innovation is a report that respondents receive after completing the survey, indicating how their network compares to the network of the average American. We hope to increase respondents' motivation to complete the entire questionnaire by the promise that they will learn about themselves. We report results from experiments with an online convenience sample, comparing the data quality achieved through the two innovations with a classical ego-centered online survey.

### **Designing an Intelligent Time Diary Instrument: Visualization, Dynamic Feedback, and Error Prevention and Mitigation**

Gregory Atkin, *University of Nebraska Lincoln*

Hariharan Arunachalam, *University of Nebraska Lincoln*

Adam Eck, *University of Nebraska Lincoln*

Leen-Kiat Soh, *University of Nebraska Lincoln*

Robert Belli, *University of Nebraska Lincoln*

Designing computer-based survey instruments requires a delicate balance between achieving the goals of the survey (e.g., collecting data, measuring response) and usability concerns from the respondent or interviewer's perspective (e.g., reducing cognitive burden and speed of use). This work presents a web-based time diary application improving upon the current computer-assisted telephone interview (CATI) instrument used in the American Time Use Survey (ATUS). We propose a novel design advancing current survey instruments through three key features based on human-computer interaction (HCI) and intelligent user interface (IUI) principles as means for improving data quality and instrument ease-of-use. First, we improve the interviewer's understanding of data in real-time using an interactive timeline which dynamically populates time diary entries while data is collected from the respondent. This design enables the interviewer to instantly assess the respondent's entire day at a glance, which (1) facilitates prompting the respondent's recall of events to avoid failures in reporting activities, (2) provides a representation of the day as a temporal sequence of activities, and (3) acts as an interview progress indicator. Second, we intelligently analyze the data as it is collected to provide dynamic feedback to the interviewer to improve interview and data quality. For example, our instrument models the respondent's day to automatically check for consistency problems, outliers, or otherwise problematic entries. The intelligent instrument also learns from past interviews to predict potential problems based on interviewer and respondent behavior (e.g., response timeliness, order of data entries). Third, using this real-time analysis, our instrument can prevent or mitigate problems during the interview by alerting interviewers of potential

problems before they occur or immediately afterward. We report on preliminary evaluations from testing the new web-based application and discuss the next steps.

### **When Proxy Interviews are Acceptable: Does it Help to Speak with the Spouse or Partner?**

Timothy Triplett, *The Urban Institute*

Doug Wissoker, *The Urban Institute*

Robert L. Santos, *The Urban Institute*

In general, proxy responses are considered less accurate than self-reported responses, but less is known about the degree of proxy reporting accuracies. Surveys accept proxy responses for three reasons. First, proxies may be used out of necessity, if the targeted individuals cannot be reached or are physically unable to respond on their own behalf. Second, proxies may be used to increase the efficiency of a survey. This occurs when you accept proxy responses for people capable of providing their own information to save time and money. Third, proxies may serve to improve the quality of the data for studies in which proxy information could provide better information than you would obtain from a direct interview.

The samples of the 2008 and 2012 Survey of Public Participation in the Arts (SPPA) allow evaluation of proxy reporting for efficiency. Within couples, respondents were selected at random and asked to report on themselves and their partners. Accepting spouse or partner proxy interviews was done as a low cost way of increasing the overall sample size. From this random selection, we have roughly equal numbers of self- and proxy reports for men and women (5,926 spouse/partner proxy reports in 2008; and 11,811 proxy reports in 2012). These surveys were sponsored by the National Endowment for the Arts and conducted as supplements to the May 2008 (n=18,444) and July 2012 (n=37,266) Current Population Survey (CPS).

This paper looks at differences between proxy and self-reported estimates for measures of attendance at cultural events. Of particular interest will be comparisons between men and women's self-reported estimates versus what their partners report they do. In addition, given the large sample size of proxy reported data we will explore whether difference in proxy reporting varies by age, education, race and possibly other population sub-groups.

### **Measurement Directness as a Cause of Bias in Reports of Socially Desirable Behavior**

Philip S. Brenner, *University of Massachusetts Boston*

Extant research comparing survey self-reports of religious behaviors, like church attendance, to time diary data have yielded evidence of extensive measurement bias. These comparisons assume that the chronological measurement procedure used in time diaries avoids the bias inherent to survey questions by measuring in a nondirective manner. That is, where conventional survey questions prompt respondents to consider the normative focal behavior, and thus invite bias, chronological measurement avoids this process by omitting explicit references to the focal behavior.

The role of directness is tested here in three studies. In each, respondents completed a conventional survey questionnaire, including questions about frequency of religious behaviors, namely religious service attendance and prayer. Respondents were then asked to participate in

a chronological data collection procedure using SMS text messaging, each differing on the directness of the measurement.

In Study 1, respondents report all changes in major activity over a period of five days, including a weekend. No activity, religious or otherwise, was specified as the focal behavior. In Study 2, respondents are asked to report specifically (and only) on religious behavior over a period of seven days. In Study 3, respondents are randomly assigned to one of two conditions, reporting either: (1) all changes in major activity with no specified focal activity, similar to Study 1, or (2) all changes in major activity with participation in religious services specifically identified as a behavior of interest.

Findings suggest that priming the respondent to consider a focal, normative behavior, like religious practice, can bias measurement of that behavior in both conventional survey questions and chronological, diary-type measurement. Moreover, respondents may change their behavior during diary data collection in response to explicit mentioning of the behavior of interest.

### **Use of Qualitative Research Techniques to Evaluate Usability of a Diary Application**

Jennifer Crafts, *Westat*

Earlayna Batch, *Westat*

Sarah Bennett-Harper, *Westat*

Jasmine Folz, *Westat*

James McClain, *National Cancer Institute*

Jana Eisentein, *National Cancer Institute*

Heather Bowles, *National Cancer Institute*

Mobile devices and applications have become increasingly attractive options for both real-time and retrospective data collection due to widespread use of smartphones. Optimizing application users' experiences is important to promote motivation and engagement with the reporting task. Developing an easy to learn, easy to use, and engaging interface as well as an intuitive task flow are especially important for applications designed for use in diary studies, where user compliance is crucial for data quality and accuracy. The focus of this poster is on the role of qualitative usability evaluation methods in the development of the US National Cancer Institute's (NCI's) My Life in a Day application. The application was designed to function as both a time use diary (to enter daily activities and events as they occur) and as a recall aid (to fill in gaps in recall or provide additional details). The application was developed as a tool for researchers to collect physical activity and related health behavior data for a broad variety of research designs, including clinical monitoring, behavioral interventions, and surveillance studies. The application features a customizable user interface and provides the ability to record daily activities with start and end timers for up to two simultaneous activities. In-person observation sessions and follow-up qualitative interviews were conducted to assess study participants' use of the application over a three-day period. We will present usability findings in terms of level of severity of issues and the effects of those issues on user perception of burden and data quality. We will also provide lessons learned for formative evaluation during application development and principles for enhancing usability of mobile applications for data collection.



## Nonresponse Bias

### Exploring of Nonresponse and Measurement Error in a Study Using Respondent Driven Sampling: Focus on Recruitment Coupon Distribution and Network Size Reports

Zeynep Tuba Suzer-Gurtekin, *ISR - University of Michigan -Program in Survey Methodology*

Respondent Driven Sampling (RDS) has been widely employed in sampling and statistical inference of rare and hidden populations. However, the underlying assumptions and practices in RDS statistical inference are not well understood by survey researchers who rely on traditional probability samples. This study focuses on the effects of two errors specific to and relevant for recruitment processes and estimation. First, nonresponse partially examined through recruitment coupon distribution appears to negatively impact the sample size growth assumed to be exponential. Second, the reports of network sizes used as an adjustment factor in RDS estimators are subject to measurement errors and appear to affect the bias properties in unknown directions and the efficiency negatively. The study uses a publicly available data, the Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATHCAP) conducted from November 2006 to August 2008.

### Application of Heckman Model in Assessing the Risk of Nonresponse Bias

Stanislav Kolenikov, *Abt SRBI*

Nonresponse problems have been at the forefront of research in survey methodology for the past two decades as response rates continue to drop. Bethlehem (2002) proposed a stochastic model for survey nonresponse in which nonresponse bias is proportional to the correlation between a survey outcome variable and response propensity. Although this framework is appealing in explaining techniques such as class adjustments to reduce nonresponse bias, the correlation in question is difficult to quantify.

We utilize a popular econometric Heckman model, joint bivariate modeling of response propensity together with the outcome of interest, to identify appropriate correlations. In the basic model, unit response propensity is fully modeled using frame variables, while the outcome equation is left in intercept-only form, which corresponds exactly with Bethlehem's formulation, and thus allows direct estimation of the correlation in question. This diagnostic application of the Heckman model allows assessing the risk of nonresponse bias, as evidenced by a significant difference between the correlation and zero. In the extended model, the outcome is modeled with frame variables, as well. We demonstrate that this extended application of Heckman model corresponds to estimation with calibrated weights, where calibration variables are used in the outcome equation.

We apply this approach to a survey of business school alumni about whom considerable information is available from the sampling frame (alumni database). The response rate (RR1) to a multi-mode survey was 25.1%, so the risk of nonresponse bias is non-negligible. The correlation between response propensity and outcome is significant in basic analyses, and is significantly reduced once calibration variables are added to the model. Thus, while unweighted analysis risks nonresponse biases, the calibrated (weighted) analysis has such risks reduced due to elimination of the correlation between the outcome and response propensity in the stochastic response model.

## **Measuring Nonresponse Bias in Web Surveys: The Role of Health Status**

Mengmeng Zhang, *University of Michigan Program in Survey Methodology*

Web surveys have been included in many national longitudinal studies and panels as an additional option offered to respondents together with, or instead of, traditional modes (e.g. CATI or CAPI). Web surveys cost less and this mode generally requires a shorter data collection time compared to the traditional modes of data collection. Evidence is mixed regarding the presence and size of coverage error and nonresponse error in Web surveys. Typically, there is little information about non-respondents to allow for the measurement of nonresponse bias.

This paper takes advantage of the 2011 Health and Retirement Study (HRS) Internet Survey to examine factors affecting people's decisions to respond to a Web survey. Respondents to the 2011 HRS Internet survey are randomly selected from respondents to the 2009 HRS internet survey and the 2010 HRS respondents who had Internet access. As a result, a lot of information is available on both respondents and non-respondents to the 2011 HRS Internet survey. Making use of the rich auxiliary data, we first characterize non-respondents to the 2011 HRS Internet survey. Specifically, we examine whether health status is a factor that might influence respondents' willingness to participate in the Web survey after controlling for age, gender, and socioeconomic status (SES). We hypothesize that respondents with lower self-rated health in previous waves were less likely to participate in the 2011 Internet survey. We then explore the role of health status in nonresponse bias in several key survey variables from the 2011 Internet survey for older adults.

## **Adjusting for Attrition Bias in a National Longitudinal Survey of Dating Violence Among Latino Youth**

Heather Hammer, *Abt SRBI*

Recent research finds that Latino adolescents are at higher risk of dating violence compared to white adolescents. Whereas Latinos are the largest minority in the U.S., they are hard-to-reach and disproportionately impacted by declining response rates and differential nonresponse. This suggests that dating violence rates for Latino adolescents may be substantially underestimated. To address this issue and examine cultural influences, the National Institute of Justice Office of Justice Programs funded the Dating Violence among Latino Adolescents (DAVILA) study. In 2011, this study conducted a CATI survey with a national sample of 1,525 Latino youth aged 12-18 (DAVILA 1). Then in 2013, DAVILA 2 surveyed the DAVILA 1 youth who were willing to do a similar survey in a year or two. Of the 1,427 willing youth who completed the DAVILA 1 survey, 574 or 40.2% completed DAVILA 2.

We found no difference in the weighted incidence of dating violence victimization between waves. We also found that DAVILA 1 dating violence victims were less likely to participate in DAVILA 2 compared to non-victims and more likely to have been dating violence victims in the 12 months prior to DAVILA 2 when they did participate.

To estimate the contribution of attrition bias and to adjust for the bias as needed, we developed logistic regression models using DAVILA 1 survey and paradata to estimate response propensity for DAVILA 2 and construct preliminary nonresponse adjusted weights. This paper will show how the R-indicators, odds ratios, and Heckman analysis results can be used to identify potential sources of attrition bias, and to refine and assess the attrition bias correction.

We will also discuss implications for the design and implementation of longitudinal CATI surveys on sensitive topics with Latino adolescents.

## **Panel Surveys: Data Quality Attrition and Retention Issues**

### **Response Effects in Panel Survey Participants: Cultural, Socioeconomic, and Life Experience Predictors**

Jennifer Benoit-Bryan, *University of Illinois, Chicago*

Allyson Holbrook, *University of Illinois, Chicago*

While response behaviors in survey participants have been studied extensively, there is much less research available that examines the changes in response behavior in individuals across panel waves. This research examines two types of response bias, extreme responding and acquiescence response bias using all four waves of the nationally representative Add Health study, which includes a sample size of more than 15,000 respondents who are tracked across fifteen years. The first wave of the study is collected using a clustered school sampling design of adolescents in grades 7-12 during the 1994-1995 school year. The most recent set of interviews occurred in 2008-2009 with a sample aged 24-32. We examine responses biases over time and assess whether consistency in response bias varies based on characteristics such as race, ethnicity, country of origin, educational attainment, and income level. For example, do recent immigrants to the United States display different patterns in response behaviors across time compared to individuals who were born in the United States? In addition, we examine the potential for other variables to affect the level of response bias over time such as drug use and mental illness. A better understanding of the consistency of response bias in individuals over time would be useful to collectors of panel survey data and help researchers better understand the underlying factors that lead to response bias.

### **Measuring Change Using Dependent Interviewing – Does It Matter How Questions Are Worded?**

Annette Jäckle, *University of Essex*

Stephanie Eckman, *Institute for Employment Research*

Tarek Al Baghal, *University of Essex*

Noah Uhrig, *University of Essex*

Emanuela Sala, *University of Milano Bicocca*

Frederick Conrad, *University of Michigan*

Panel surveys are used to measure changes experienced by individuals over time. The extent of change is however usually exaggerated due to reporting errors. Most panel surveys therefore use Dependent Interviewing (DI), which has been shown to reduce spurious change. With the more common form of proactive DI, respondents are reminded of responses given in their previous interview, before being asked about their current situation. DI questions can be worded in different ways. After reminding a respondent of her report in the last interview (“Last time we interviewed you, you said you were employed.”), we might ask yes/no questions: “Is that still the case?” or “Has that changed?”; or forced choice questions: “Is this still the same or has it changed?”; or we might ask the original question again: “What is your current activity?”. All of these different wordings are currently in use, sometimes within a single survey. The literature provides no guidance about how to choose between different wordings.

We use data from two experimental studies using objective/subjective and categorical/continuous variables (the CAPI Innovation Panel of the UK Household Longitudinal Study and a CATI survey in Germany) to examine whether the wording of DI questions matters; through which mechanisms the wording matters; and which wording produces more accurate measures of change. Initial results suggest that there are some differences between DI wordings. These differences are not associated with indicators of personality, acquiescence tendency, cognitive ability, effort, or distraction during the interview. In the CAPI experiment, differences between DI wordings seem to be associated with interviewer behaviours. For the CATI experiment, we will in spring 2014 be able to link to administrative records on labour market activity and income sources. We will use the linked record data to assess which DI wording produces the most accurate measures of change.

### **Survey Breakoff in Online Panels**

Allan L. McCutcheon, *University of Nebraska-Lincoln, Survey Research and Methodology*

Although survey breakoff appears to be somewhat less likely in online panels than in usual web surveys – due, at least in part, to early panel attrition by less-committed panel respondents (see e.g., McCutcheon, Rao, and Kaminska, forthcoming) -- breakoff in online panels remains problematic. The increased interest in breakoff involving internet survey respondents has been accelerated by the relatively recent availability of paradata collection methods for web surveys (Peytchev 2009). In addition to respondent and survey design characteristics, it is now relatively easy to obtain data such as the amount of time taken per survey item (response latency), number of response changes, time of day that survey breakoff occurs, and other factors that may contribute to survey breakoff. Internet panels offer the additional advantage of accumulated information regarding respondent characteristics.

This study examines data from multiple waves of the internet component of the Gallup Panel, a multi-mode, probability panel of American households. In addition to demographic characteristics and survey design factors (e.g., question complexity, topic, number of questions, survey length), the analysis will include self-reports on internet sophistication, paradata and device type to explore factors related to survey breakoff. Preliminary analysis indicates that while long-term panel members are less likely to breakoff, past behavior is an important predictor – those who have broken off in the past appear to be more likely to do so again. Also there appears to be a clear and persistent pattern with respect to response latency; as respondents approach breaking off their survey participation, they tend to slow down in their response time (increase response latency). The study will explore the potential use of such predictive models for survey breakoff in designing possible adaptive design (Groves and Heeringa 2006) interventions for internet surveys that may prove useful in delaying survey breakoff.

### **Correlates of Attrition in the German Internet Panel: Drop-Outs and Sleepers**

Annelies G. Blom, *University of Mannheim*

Gabriele Durrant, *University of Southampton*

Which characteristics are associated with continued participation in a longitudinal survey? Answering this question is essential in any longitudinal survey and has been well-researched for face-to-face and telephone surveys. However, in the context of probability-based online surveys little is currently known about the correlates of attrition.

This paper examines the correlates of attrition and nonresponse in the German Internet Panel (GIP) using longitudinal and multilevel data analysis methods for clustered and repeated measures data. The GIP is based on a true probability sample of individuals and was the first of its kind to be conducted in Germany. In 2012 the recruitment of the GIP was first carried out offline through face-to-face interviews. To minimize coverage- and nonresponse errors households without access to the internet were equipped with the necessary hardware and/or a broadband internet connection. After online registration interviewing takes place bi-monthly on topics of political and economic behavior and attitudes. By May 2014 the GIP will have conducted ten waves of data collection.

We first report results from the sequence analysis to better understand attrition patterns across all waves. Then we model participation at the first wave (recruitment stage) and nonresponse and attrition in subsequent waves. Our modelling approach aims to take account of the clustering of individuals within households and interviewers (from the initial recruitment stage) and of responses across waves (repeated measures). A range of covariates are available including both time-invariant and time-varying factors. We specifically focus on what we may learn from the paradata collected online at each wave, such as response times, break-offs and (mobile) device used. Initial analyses show various patterns of participation across waves, including respondents who drop out completely and sleepers who return to the panel. Our work will have implications for panel maintenance measures in online probability panels.

**Thursday, May 15**  
**3:00 p.m. – 4:00 p.m.**  
**Poster Session 1**

**1. American Catholics: A Latent Class Analysis**

Ann M. Arthur, *Gallup, Inc., University of Nebraska-Lincoln*

Allan L. McCutcheon, *Gallup, Inc., University of Nebraska-Lincoln*

Catholicism is one of the largest religions worldwide and the largest religious minority in the United States. While Catholics have united as a cohesive voting bloc in the past, declining religiosity and increasing secularization has splintered American Catholics into homogeneous subgroups. Three subgroups are commonly recognized: Conservative Catholics, who adhere closely to the religion, "Cafeteria" Catholics, who pick and choose aspects to believe and follow (Dobbelaere, 1981), and Non-Practicing Catholics, who self-identify as Catholic but otherwise do not practice the religion.

Several researchers note that religiosity is not necessarily unidimensional (e.g., Glock, 1962; O'Connell, 1975), and consists of at least five dimensions: ideology (beliefs), ritualism (practices), experiences (emotions and feelings), intellect (knowledge), and consequences (prescriptions for daily living that result from the religion) (Glock, 1962). This research examines the religiosity of American Catholics in a broad sense using five indicators of religiosity: belief in God, strength of religious affiliation, frequency of church attendance, frequency of prayer, and the extent to which one considers himself or herself to be religious.

This research uses latent class analysis to quantitatively characterize American Catholics into subgroups. Data for this research come from the 2010 General Social Survey and consist of a subset of 1,113 Americans who self-identify as Catholic. Preliminary analyses indicate that there are actually four latent classes of American Catholics – the three previously hypothesized subgroups (Conservative Catholics, 30.5%; "Cafeteria" Catholics, 41.9%; Non-Practicing Catholics, 21.3%) as well as a fourth group of Ritualists (6.4%), who engage in religious practices despite not believing in God with certainty. In this paper, the four latent classes are identified and described in terms of characteristics and size. Additional analyses are conducted to further describe membership in the classes. The results of the research are then discussed in the context of the United States and worldwide trends.

**2. Dealing with Deaths in Longitudinal Surveys**

Nicole Watson, *University of Melbourne*

Longitudinal surveys follow people over time and some deaths will occur during the life of the panel. Through fieldwork efforts, some deaths will be known but others will go unobserved due to sample members no longer being issued to field or having inconclusive fieldwork outcomes (such as a non-contact not followed by a contact at a later wave). The coverage of deaths identified amongst sample members has flow-on implications to nonresponse correction. Using the Household, Income and Labour Dynamics in Australia (HILDA) Survey, we examine the extent of missing death reports through two methods. The first method extrapolates the expected number of deaths in the original sample (selected in 2001) over the first 11 years using life expectancy tables. The second method matches the sample to the National Death Index. Both methods have challenges in their implementation

and these are explored. We further examine the impact of missing deaths on assumptions in the construction of weights for the balanced panel and subsequent population inference.

### **3. Assessing Nonresponse and Panel Conditioning in an Intensive Measurement Design**

Carlos Macuada, *University of Michigan*

Megan E. Patrick, *Institute for Social Research*

Jamie Griffin, *Institute for Social Research*

Panel studies allow researchers to analyze a sample over time. However, this method may have a direct impact on respondents' attitudes and behaviors (Toh, Lee & Hu, 2006; Tourangeau, Rips & Rasinski, 2000). Panel conditioning, also known as measurement reactivity (Barta, Tennen & Litt, 2012), is understood as the change in someone's behavior as a result of frequent measurements (Das, Toepoel & Soest, 2011). Although this effect was identified more than 5 decades ago (Campbell, 1957), the literature lacks an applied framework by which to measure and disentangle it from other error sources, such as attrition bias (Warren and Halpern-Manner, 2012).

The survey used here followed a sample of high school seniors during a twelve month period. This particular research focused on 202 students who were assigned to an intensive measurement, or measurement burst, design (Sliwinski, 2008). After an initial baseline in the spring of 2012, students were re-contacted four, eight, and twelve months later to complete fourteen brief daily web-based surveys each time. The substantive focus of the research was substance use and related risk behaviors.

To separate possible panel conditioning effects from panel attrition effects, we first examined substance use patterns across all days reported, and then examined patterns only among those who responded to all 14 days (to isolate the potential conditioning effect). There was limited evidence for any trend in reports of substance use across weeks. Across all sampled days, there was a decrease in marijuana use (during Wave 1 only). After restricting the comparison to complete responders, there were no differences. Results suggest that any differences observed in this intensive measurement design are likely due to nonresponse across the 14 days, rather than panel conditioning.

### **4. Are Nonresponse Error and Measurement Error Related? Evidence Using a Panel Study**

Peter Lugtig, *Department of Methods and Statistics - Utrecht University*

Survey methodologists worry that errors due to nonresponse and measurement interact. Some reasons for nonresponse might at the same time also be a reason for reporting with more measurement error. Lower cognitive abilities, lack of motivation, or language difficulties are among the possible reasons for both nonresponse and measurement error in surveys.

Studying the interaction between both survey errors is complicated, because by definition we have no survey data for nonrespondents, and thus cannot estimate measurement errors for them. Often, methodologists have looked at measurement errors for both 'early' and 'late' respondents as a proxy for understanding whether there is any tradeoff (e.g. Cannell and Fowler 1963; Kaminska and MCutcheon 2010).

In this paper, the trade-off between measurement error and nonresponse error is investigated for respondents in the British Household Panel Survey. Using data from 18 waves, attrition profiles are created using Latent Class Analysis, based on the timing and nature of attrition. For each of the attrition profiles, we then see how much measurement errors respondents reported with before they dropped out of the panel survey. We separately look at the reliability of survey data using repeated measures of the same variable, and the validity of scales at cross-sectional waves of data.

We find that measurement errors and nonresponse errors indeed interact, and that respondents who drop out early in the panel survey report with more measurement error before dropout, than respondents who drop out late in the survey, or those who continue to be loyal panel members. We conclude with a discussion how nonresponse follow-up strategies in a panel survey should be evaluated in light of our findings.

## **5. The Effect of User Experience (UX) Design on Data Quality: A Re-Design of an Online Diary**

Yelena Pens, *Nielsen Company*

Robin Gentry, *Nielsen Company*

How do you transform a traditional paper-based instrument that has been in use for 60+ years into a modern web-based application? This paper will share the results an iterative testing process of usability studies and general population tests to develop a web-based one-week radio diary and show how the results compare to the traditional paper diary. These results show the advantage of incorporating user experience testing into the testing process.

The first web application, or eDiary, pilot test was conducted in 2011. The results of the test identified potential usability issues with the online application such as overlapping time entries and conflicting time issues. Following the field test, a usability study of the online diary application was conducted. Web savvy and non-web savvy participants were invited to use the application and provide feedback on the user experience. The focus of the usability study included content and features of the online diary such as radio buttons, checkboxes, and a slider bar. In winter 2012, an in-home ethnography study was conducted in three markets where participants used scenarios and completed tasks to enter radio listening entries in a diary page. Additionally, expert reviewers were commissioned to review the usability of the overall website from registration to completion of diary keeping. The results from the three studies were consistent and resulted in a re-design of the eDiary application. A second pilot was conducted in 2013, and the results of the second pilot were encouraging. We will present the results from the eDiary usability study, and the comparison of the results from the pilot tests. We will show the features of the online diary and the effect UX had on data quality such as time precision and duplicate entries. The results show that a simple approach is usually the optimal one.



## **6. A Lesson in Questionnaire Design: Updated Images, Motivational Messages and Unintended Consequences**

Christine Pierce, *Nielsen Company*

Lauren Walton, *Nielsen Company*

Anh Thu Burks, *Nielsen Company*

Survey researchers often apply visual design and motivational messages in an effort to increase response rates in self-administered surveys. When effective, graphical images and persuasive communications will appeal to respondents and have a positive influence on their response propensity. This research explores the results from a redesign of the Nielsen TV Diary materials and the impacts on respondent participation. In an attempt to motivate respondents, particularly hard-to-reach younger and ethnic populations, the redesigned TV Diary materials focused on the importance of participation, tailored images, and updated graphics. The test design compares the existing TV Diary (control) and the redesigned TV Diary (test). The key comparative success measure is the cooperation rate – defined as usable returned diaries as a proportion of total mailed diaries.

Despite pretesting to a series of focus groups, the redesigned TV Diary showed lower overall cooperation rates. Results were statistically significant (18.4% for control vs. 16.7% for test). As articulated by the leverage-salience theory (Groves et al 2000), the effectiveness of a given survey attribute will vary for different respondents. Similarly, the study reveals that designs intended to increase the participation of hard-to-reach groups may have the opposite effect on other members of the respondent pool. This analysis uses geo-demographic tools to highlight notable response patterns and to demonstrate the varying levels of effectiveness for different population segments. Lastly, the paper highlights how qualitative pretesting may fall short of anticipating impacts in the general population.

## **7. Sequential or Simultaneous Multi-Mode? Results from Two Large Surveys of Electric Utility Consumers**

Carla Jackson, *Abt SRBI, Inc.*

Christine Ledoux, *Southern Company*

A recent meta-analysis of the response rates in 19 multi-modal surveys offering mail and web options indicated that a concurrent web option significantly lowered overall response rates (Medway and Fulton, POQ, Winter 2012).

The impact of sequential or simultaneous multi-modal survey approaches was analyzed by comparing results from the two most recent waves of a residential appliance saturation survey conducted for Southern Company.

The survey, which is conducted at three-year intervals, was conducted in 2010, and data collection for the 2013 survey wave has just been completed, with samples of 15,000 randomly-selected customers in each wave. The following data collection approaches were used in the two survey waves:

- Simultaneous multi-mode: In 2010, the cover letters transmitting the original mail survey provided a URL and password at which the potential respondent could access the survey online, rather than returning it by mail. This information was also included in a second mailing of the questionnaire to non-respondents.

- **Mixed sequential/simultaneous multi-mode:** In 2013, the survey began with an email invitations to approximately one-half of respondents for whom email addresses were available. A reminder email was sent several days later. Subsequently, mail survey packets were sent to all non-respondents, including both those who had and had not received the email invitation. The URL and a password to access the survey were included in both the initial and reminder cover letters for the mail survey.

There was an increase in the number of web surveys completed in 2013 compared with 2010 (703 and 476, respectively). However, the overall response rate decreased from 36.9% in 2010 to 29.9% in 2013. We are currently analyzing 2010 and 2013 response rates and information from 30 quality control checks

## **8. Difficult Data: Comparing the Quality of Behavioral, Recall, and Proxy Data across Survey Modes**

Oana M. Dan, *Nielsen Company*

Kumar Rao, *Nielsen Company*

Vera Kurmlavage, *Nielsen Company*

The mode choice literature is rife with evidence on the impact of different survey modes on response rates, respondent cooperation, and data quality. However, insufficient attention has been paid to the quality of “difficult data” provided when respondents cannot choose the mode and thus cannot maximize their comfort with the survey. Here, “difficult data” correspond to questions that are burdensome for respondents to think about – e.g., very specific details on a behavior, on past events, or on the behavior of other persons. Which mode best helps respondents provide difficult data with fewer break-offs, straight-lined responses, and “I don’t know’s”? Do cues from a phone interviewer help respondents recall elusive details? By contrast, does the simultaneous interaction with the interviewer intimidate respondents and constrain their ability to think through the answer? Similarly, does pacing oneself through an online survey provide respondents the time and comfort needed to think carefully through the responses, or does it weaken the motivation to complete the survey altogether? This paper assesses the impact of survey mode (without choice) on the quality of “difficult data,” focusing on a survey that was administered via CATI and online. The survey collected very specific information on media consumption on mobile devices by all persons in a household, as reported by one “proxy” respondent. The CATI survey was administered on a dual-frame probability sample, and the online survey was administered on an opt-in panel of mobile users. While the CATI sample is representative, the mobile user sample was better positioned to provide the information required, allowing for insightful comparisons and directional evidence on mode effect on the quality of difficult data. The paper both presents findings from these surveys and raises a call for further research on survey mode, difficult data, and respondents’ place between them.

## **9. Using Vendor Appended Data in a Two-Stage Address Based Sampling Design for Cost Savings**

Kelly Dixon, *Nielsen Company*

Mike Kwanisai, *Nielsen Company*

Alan Tupek, *Nielsen Company*

Researchers in all disciplines are developing approaches to improve sample design by appending external data to a frame. Nielsen uses a two-stage address-based sample design to recruit people to participate in a media ratings panel survey. During the first stage, key

demographic information is gathered about the household, and in the second stage individuals are recruited into the panel. Households designated as replacement households only progress to the second stage if replacement sample is needed. Replacement sample is more likely needed for demographic groups with typically poor participation rates in panel surveys.

In our study, we used the vendor-appended data to select potential Hispanic and young adult replacement households for the first stage of selection at higher rates than other households. This approach provides a pool of replacement sample that is richer in demographic groups that need replacement sample. Our results compare the demographic information from the vendor to the replacement sample at each stage of selection. We discuss success metrics in terms of accuracy and coverage and the importance of each to make the two-stage process more efficient and cost effective.

## **10. Considerations for and Lessons Learned from Online, Synchronous Focus Groups**

Sarah Forrestal, *Mathematica Policy Research*

Angela Valdovinos D'Angelo, *Mathematica Policy Research*

Lisa Klein Vogel, *Mathematica Policy Research*

Tessa Kieffer, *Mathematica Policy Research*

Nyna Williams, *Mathematica Policy Research*

Advances in technology continue to push the boundaries of collaboration, bringing together geographically dispersed people into a single space through easy-to-use virtual tools. The availability of such technology, combined with pressures to minimize data collection costs, has encouraged the use of online, synchronous focus groups. Despite these trends, little literature exists on the use of online focus groups. In this paper, we describe considerations for when to use online focus groups as well as lessons learned from implementing them during two recent studies of hard-to-reach professional populations.

We considered several factors when we chose to use online focus groups with U.S.-based clinicians and internationally-based IT professionals. First, participants' geographic dispersion made the traditional in-person format impractical. Second, participants must have reliable Internet access and be comfortable with technology. Third, the relative anonymity allows participants to provide candid responses to questions. Finally, costs are reduced by eliminating travel and facility fees, and incentives can be lower because of greater convenience for participants. However, staff costs may be greater for preparing presentation materials or setting up the technology.

We learned several lessons for future online focus groups. "No-shows" are at least as likely for online groups, and overscheduling should be used as it would for in-person groups. To encourage engagement, small groups are preferable, and moderators should explicitly ask participants not to multi-task. Checking in with the whole group frequently through polls or round-robins is an effective engagement strategy because rapport is harder to establish in the absence of nonverbal cues. To address the lack of nonverbal cues, video may be preferable to an audio-only format if the moderator and all participants have the required technology. Data collection team dress rehearsals can minimize time wasted due to technical problems.

## **11. From Data Sharing to Data Stewardship: Meeting Data Sharing Requirements Now and into the Future**

Linda Detterman, *Institute for Social Research - University of Michigan*

There exists a growing desire, and growing requirements for scientific research data collected by federal funds to be shared publicly and without charge. Agencies such as the NSF and NIH require data management plans as part of research proposals and the Office of Science and Technology Policy (OSTP) is requiring federal agencies to develop plans to increase public access to results of federally funded scientific research.

To be effectively shared, data must be described and documented, discoverable online, and accessible, today and into the future. And sharing data into the future requires that data sharing entities are sustainable. Data sharing and storage, as well as discoverability and accessibility, is not free from costs. Sustainability requires funding.

ICPSR, a center within the Institute for Social Research at The University of Michigan, has been sharing and archiving social and behavioral research data for over 50 years. This poster will explore several data management models including:

- Fee for access model - pooled funding for data curation and preservation for access by the pooling members
- Agency-funded model – agency or foundation funded model providing free public access
- Fee for deposit model – fee for deposit of data to provide free public access

The poster will also cover restricted-use data sharing in the public access environment.

This poster is useful to research scientists and those working with scientists who are interested and required to share their research data. Tips for accessing public data access products will be provided as well as resources for creating data management plans for grant applications.

ICPSR views data sharing as “data curation” akin to work performed by an art or museum curator. Through data curation, we enable today’s research scientists to be recognized and cited, and tomorrow’s scientists to maximize the benefit of research investments.

## **12. Can Visual Design and A Verbal Importance Prompt Reduce Item Nonresponse For Demographic Items?**

Glenn D. Israel, *University of Florida*

Conventional wisdom and practice lead to placing demographic items at the end of a questionnaire. The thinking behind this practice is that these items are less important than topically-salient items for most surveys, so higher item nonresponse can be tolerated for demographic questions. A recent study by Teclaw, Price and Osatuke (2012) turn this logic on its head and found that item response for demographic items at the beginning of a questionnaire was higher than for the same set of items at the end of the survey. This finding raises the question of whether there are other equally effective approaches to stimulating high item response rates for demographic questions. This study reports on two experiments testing whether visual design and a verbal prompt about the importance of answering the demographic questions improves item response rates when the items are placed at the end of the survey. Data from a customer satisfaction survey of Cooperative

Extension Service clients are used to address the research question. The mixed-mode survey data included both Web and mail survey responses. Overall, the item response rate was no higher for the questionnaire with the verbal prompt than the one without it and the version employing visual design to emphasize the prompt resulted in a non-significant lower item response. In addition, item response rates were not different for either the mail or Web responses (although the later showed a higher item response rate with the prompt, it was not statistically significant). Based on these results, it does not appear that visual design and verbal prompts about importance is a viable strategy for reducing item nonresponse of demographic items.

### **13. Computing Survey Response Rates for Probability-Based Web Panels Recruited Through Multiple Frames and Modes Across Multiple Cohorts**

Charles DiSogra, *Abt/SRBI*

Mario Callegaro, *Google UK, Ltd.*

More and more probability-based web panels have been created, especially in Europe, or are being planned in the U.S. Surveys conducted with these panels require an accurate calculation of response rates. Computing accurate response rates is important to assess nonresponse bias when samples are drawn from these panels or when the entire panel is used for a single survey. Online panels can be classified in two distinct groups: “volunteer” or non-probability access panels, and pre-recruited probability-based panels. The volunteer panels have no recruitment sample frame (members opt-in pro-actively) and thus the probability of selection is unknown. For these non-probability panels the computation of response rates is not possible since there is no defined denominator. In the second group, computing response rates for the probability-based panels is possible by taking into account the different stages of panel creation and each stage’s related response. The computation can be complicated and can also vary in its functional components due to different recruitment methods used by panels and how each panel manages their membership. This presentation addresses the recent growth of mixed mode, multi-frame panel recruitment methods (necessitated by current cost-efficiency strategies) and how to handle their response rate computation. Building upon earlier-published and somewhat simpler models, it is now possible to construct multifactorial formulae to compute response metrics for surveys derived from probability-based web panel samples accounting for the complexities of multiple-recruitment cohorts, mixes of recruitment modes, multi-frame recruitment and several dimensions of panel life dynamics to arrive at a true response rate. Incorporating the multiplicity of factors involved affecting various recruitment cohorts, typical “heterogeneous” samples drawn from these panels require a weighted response rate to be computed. Examples of these computational models and weighted response rates will be presented and discussed including response rates applicable to longitudinal studies.

### **14. Nonresponse Bias and Mode Effects in a Survey of Fishing Effort**

Marci Schalk, *Abt SRBI*

Courtney Kennedy, *Abt SRBI*

Sujata Pal, *Abt SRBI*

Alex Shapiro, *Abt SRBI*

Rob Andrews, *NOAA*

Many survey researchers are exploring ABS mail surveys, but the error properties of these surveys are not fully understood – especially as they compare to dual frame RDD surveys. While mail surveys have some advantages over CATI surveys (e.g., self-administration),

they may be more susceptible to nonresponse bias. The fact that mail respondents can peruse the entire questionnaire before deciding to participate, potentially makes it more likely (relative to CATI respondents) that they will base their decision on the survey topic. For surveys such as the 2012 NOAA Recreational Fishing Survey, the risk is that anglers will be more likely to respond than non-anglers – and that this differential may be compounded by the use of mail mode.

In order to test this and related hypotheses, the Recreational Fishing Survey was designed with randomized mode (CATI/mail) experiments. This multi-wave study was conducted in two phases. To identify recreational saltwater anglers in the general population, a Screener survey was sent by mail to an ABS sample of residential addresses. In the second phase, a multi-mode Angler survey was conducted using two sample sources: (1) adult anglers identified from households who completed the Screener survey, and (2) a list sample of individuals with saltwater fishing licenses in the study area.

This presentation examines nonresponse bias and mode effects by analyzing two experiments. In the Angler survey, individuals with a known telephone number were randomly assigned to either mail or CATI mode. A Nonresponse Follow-up Study (NRFU) was conducted of households that did not respond to the Screener survey and also featured a randomization of survey mode. The results from these experiments provide guidance on the presence and magnitude of mode effects in reports of fishing behavior, as well as the extent to which differences in estimates may be attributable to nonresponse bias.

## **15. Motivation Research: Combining Qualitative and Quantitative Methods to Understand the Drivers of Behavior**

Julie Paasche, *Artemis Strategy Group*

Anne Aldrich, *Artemis Strategy Group*

Dave Richardson, *Artemis Strategy Group*

Our goal as researchers is often ultimately to promote policies, position brands, or assist with the launch of new products and services.

Understanding what motivates people to make specific decisions is critical to success. Based on the means-end theory, motivation research explores how the tangible aspects of a product or service connect with the emotional needs of the audiences. Motivating consumers, customers, supporters or voters to act requires a feat of persuasion that combines the rational and repeatable building blocks of behavior with the emotional and visceral appeals to our core nature. When we think about any initiative to get people to act in desirable ways – buy, support, vote – there usually are many ingredients in the process. If the initiative relates to a product or service, the tangible value of the product/service attributes and the support behind them are obviously important. If the desired change relates to an issue, conveying just the tangibles often does not provoke a decision to act. People ultimately need to discern the personal relevance of the decision through some level of persuasion. The insights gained through motivation research provide a strategic template for shaping persuasive communications strategy and messages to strengthen the personal relevance of a concept, mission or brand.

This presentation covers means-end theory, and the use of a two-step research process: a qualitative phase built around in-depth individual and small group interviews to probe the key tangible and emotional benefits and underlying values, and a quantitative phase

focused on target populations to understand sizes and demographic characteristics of groups that follow similar pathways of thought. This analysis is the foundation for building persuasive messages.

The presentation will conclude with several cases studies from areas such as financial management, public policy, voting behavior, and food purchases.

## **16. The Use of Multiple Methods to Evaluate Response Problems**

Jennifer Beck, *U.S. Census Bureau*

The Commodity Flow Survey (CFS) provides the only data on the shipment of goods on United States roadways. The U.S. Census Bureau conducts the CFS every five years, on behalf of the Bureau of Transportation Statistics (BTS), as part of the Economic Census. For the CFS, businesses provide quarterly information on the type, destination, and value of the goods shipped during a predetermined week of the quarter. In an effort to improve data collection procedures for future survey cycles and to evaluate previously identified issues with the response process, the Census Bureau conducted an evaluation study of the CFS. Rather than limit the evaluation to a single method, we first analyzed collected data to identify issues that caused the most problems for data collection and processing. Then, using that information, we identified groups of respondents who experienced those issues to target for in-depth, follow-up interviews. Through this multi-method approach, we were able to obtain focused respondent feedback and develop recommendations for future improvements to the CFS. We discuss these results and highlight the efficiency of using this multi-method approach as a tool for evaluating data collection procedures and sources of survey error.

## **17. An Examination of Opposing Responses on Duplicated Multi-Mode Survey Responses**

Amy L. Djangali, *IMPAQ International, LLC*

IMPAQ International regularly conducts surveys to support various government evaluative research as well as stand-alone survey projects. These surveys include a multi-mode approach which may result in duplicate cases across waves or modes of the data collection. In these instances, typically only one case is retained, which is selected based on fairly simple rules such as the most complete or the first to arrive. However, in reviewing the actual response content of several duplicates for multi-mode surveys, we noticed that the respondents who sent in two or more surveys often provided different answers to the same question. While it can be common for survey respondents to move one degree in subsequent responses (e.g. from strongly agree to agree), we observed that many respondents moved two or more degrees (e.g. from strongly agree to disagree) or provided opposing responses (e.g. from a yes to a no on the same question). These discrepancies cannot necessarily be attributed to an extended timeline as the time between each survey receipt is short (10 days or less in most cases).

Our presentation will look at this issue across multi-mode surveys and will seek to provide insights as to why this type of discrepancy occurs. In addition, we will also examine the impact of survey mode and the use of multiple waves and how it may affect the number of disparate responses observed in the data.

## **18. Web Today, Mail Tomorrow: Mode Choice in a Longitudinal Survey**

Melissa Krakowiecki, *Mathematica Policy Research*

Larry Vittoriano, *Mathematica Policy Research*

Matt Potts, *Mathematica Policy Research*

Karen CyBulski, *Mathematica Policy Research*

Cathie E. Alderks, *Center for Behavioral Health Statistics and Quality SAMHSA*

The National Survey of Substance Abuse Treatment Services (N-SSATS) is an annual multi-mode establishment survey conducted by Mathematica Policy Research on behalf of the Substance Abuse and Mental Health Services Administration (SAMHSA). The survey includes information on facility characteristics, treatment services provided and client count information for over 17,000 substance abuse treatment facilities across the nation. The survey also provides data for the National Directory of Drug and Alcohol Abuse Treatment Programs. While facilities have the option of completing by web, mail or telephone, in any year, a majority of the facilities participate via the web. To encourage completion by web almost half of the data requested is pre-filled with the information provided from the prior year's completed survey, regardless of the prior mode of completion. Though the number of facilities that complete on the web increases on a yearly basis, there are some facilities that complete their survey in another mode. To better understand why a facility may choose to not participate via web we conducted a qualitative analysis of facilities that responded to the 2012 or 2013 N-SSATS via the mail. There are three types of respondents in the mail mode: always mail, sometimes mail but never web, and sometimes mail and sometimes web. The facilities in the analysis were ones in the third group that completed the 2012 or 2013 N-SSATS by mail and had completed one of two prior rounds via web. In addition, we divided the facilities by whether or not the respondent was the same for each round. If the facility completed via web at least once during either of those two years, they were sent an email requesting feedback regarding their choice of mode of completion. This paper is a descriptive analysis of those emails, as well as the plan for moving forward.

## **19. Using Adaptive Design to Increase Response Rates in an At-Risk, Youth Population**

Lisbeth Goble, *Mathematica Policy Research*

Jillian Stein, *Mathematica Policy Research*

Felicia Hurwitz, *Mathematica Policy Research*

Lisa Schwartz, *Mathematica Policy Research*

Obtaining high survey response rates from youth populations present unique challenges for longitudinal survey data collection. Young people tend to be highly mobile, have a limited "electronic footprint", live in cell phone-only households and use non-traditional modes of communication. These challenges are exacerbated when focusing on a low-income, at-risk population of youth such as those that participate in YouthBuild programs. YouthBuild provides at-risk youth, primarily high school dropouts, with education and counseling services as well as job training skills. Mathematica, along with its partners MDRC and SPR, are currently conducting a seven year evaluation to assess the effectiveness of the program, which includes youth surveys at 12, 30 and 48 months after random assignment into a YouthBuild program. The evaluation is funded by the Department of Labor, Employment and Training Administration.

To address the challenges of data collection with this population, we are using an adaptive design approach to increase response rates (Groves and Heeringa, 2006). This approach



uses survey and paradata to identify the most effective ways to tailor locating and contacting methods in subsequent rounds of data collection.

In this paper, we use survey data from our baseline and 12-month data collections along with paradata from the 12-month follow-up survey to examine factors related to completion of the follow-up survey. These factors include respondent characteristics (such as age, race, sex), the availability of contact information (such as phone, address, email), and the types of contact strategies used during the 12-month survey follow-up (such as the number and mode of interviewer contacts, intensity of locating effort, use of social media). Understanding the factors that increase participation will help us realize efficiencies in subsequent rounds of data collection and may provide insight.

## **20. Experiments with Email Formatting**

Benjamin Phillips, *Abt SRBI*

Stephanie Lawrence, *Abt SRBI*

Email is essential to the success of many web surveys, facilitating easy completion. Despite the importance of email to web surveys, the nature of email invitations has received surprisingly little attention. We suggest that extensive formatting lowers response rates and recommend using minimal formatting, comparable to a typical person to person email, due to display problems and resemblance to bulk email. We present the results of random assignment experiments on the impact of email formatting on survey response, conducted on elite populations.

Experiment 1 tested two conditions in an email reminder: extensive and elegant HTML formatting against minimal HTML formatting (a scanned signature was included but otherwise the email appeared as text-only). The minimal formatting condition yielded more responses.

Experiments 2 and 3 will be fielded in late 2013. Experiment 2 replicates Experiment 1 in an email invitation (as opposed to a reminder), comparing extensive HTML formatting against minimal HTML formatting. We repeat the treatment conditions of Experiment 1 to determine whether the results hold true for invitations. It is at least arguable that an extensively formatted invitation serves to enhance the legitimacy of the request by emphasizing the authority of the requestor and may enhance response to later reminders. Experiment 3 extends to the logic of Experiment 1 by comparing the minimal formatting used previously against Outlook-style formatting used by the sender in ordinary person to person email.

Experiment 4 used a two-way crossed design. Based on marketing literature, the treatment conditions varied mentioning the subject's name in the subject line in the email vs. not and including a photo of the sender vs. not. Consistent with our hypothesis, features inconsistent with ordinary email (photo and name mentioned in subject line) had fewer responses. Interestingly, personalization of the subject line generated higher open rates but fewer completions.

## **21. Surveying Community Stakeholders: Exploring Methods and Sharing Findings**

Barbara J. Robles, *Board of Governors of the Federal Reserve System*

Community stakeholders are opinion makers and leaders knowledgeable about community economic conditions. Community economic development stakeholder surveys are often

administered only at the local level by grant-funded foundation-university-public sector partners. Sometimes these community surveys are called 'asset-mapping' or 'community needs assessments' and focus on limited respondent pools. One of the ways in which local economic conditions occurring at the community-level can be captured for a more aggregated perspective is to identify regional community stakeholders and administer a national survey on community economic development conditions. We present 12 quarters of survey findings at the regional level as well as at the national level. Survey results include trend information on: workforce development issues, affordable rental housing, and top ranked current and emerging community conditions facing non-profit organizations serving low-to-moderate income communities.

## **22. Examining the Relationship Between Measurement and Nonresponse Error in a Two-Phase Survey**

Jonathan Mendelson, *Fors Marsh Group*

Luciano Viera Jr., *Fors Marsh Group*

As discussed by Tourangeau, Groves, & Redline (2010), despite longstanding interest by survey researchers, there is surprisingly little evidence of a link between nonresponse and measurement error. Indeed, a recent National Academy of Sciences report on nonresponse (2013) indicated that "research is needed on the impact that reduction of survey nonresponse would have on other error sources, such as measurement error." One reason for the difficulty of demonstrating this link is that measurement error can typically only be measured for survey respondents, not for nonrespondents. Recontact studies circumvent this problem, offering a unique opportunity to investigate this possible link by examining measurement error in the initial phase in conjunction with nonresponse in the recontact phase.

This paper assesses the relationship between measurement error and unit nonresponse in a two-phase study sponsored by the U.S. Department of Defense. The initial phase consisted of a national mail survey of young adults ages 16–24, sampled from an address list database covering over 90% of the target population, assessing attitudes and behaviors pertaining to military recruiting. Respondents were subsequently recontacted for a follow-up survey, completed online, assessing the awareness of and attitudes toward the Military's advertising campaigns.

We examine the relationship between data quality in the first phase and unit nonresponse in the recontact phase. Indicators of data quality include item nonresponse, lack of response differentiation for grid items, and response inconsistency. Preliminary analyses demonstrate a link between item nonresponse and subsequent unit nonresponse; for example, participants who did not answer a question in the initial study regarding educational intentions completed the follow-up phase at a significantly lower rate (31%) than did participants who did provide an answer (40%),  $p < .001$ ; similar results are attained for other questions. After assessing the relationship between measurement and nonresponse error, we discuss implications for survey practice.

### **23. Risky Business: Can Visual Cues and Mental Exercises Affect a Person's Decision-Making Process for Engaging in Risky Behaviors?**

Morgan S. Jones, *East Tennessee State University*

Reagan Barbee, *East Tennessee State University*

Kelly N. Foster, *East Tennessee State University*

New advances in technology afford us the opportunity to engage survey participants in new ways using graphic images, audio and video, and interactive vignettes. The purpose of this study is to investigate the impact that a strategic visual cue and mental exercise has on a person's decision-making process, their likelihood of engaging in risky behaviors, and how those behaviors can impact future outcomes.

This project surveys students at a major state university on a series of demographic and decision questions intended to measure their how likely they are to engage in risky behaviors. Measurements will consist of a baseline assessment with a follow-up 7 days later. At time point 2, students are randomized into one of two conditions. The treatment condition shows students an age-progressed version of their university picture and then engages them in a 10 year planning exercise. They are asked what they think they need to be doing 5 years from now, 1 year from now, and within the next month to meet their personal, educational, and career goals followed by replication of the baseline decision questions. The control condition shows students their current university ID photo, a university welcome video, and the baseline decision questions.

The most significant contribution of this research is that there is a critical need to develop a cost-effective, and widely implementable intervention that causes students to permanently change their attitudes about how present-day choices about risky behaviors affect their future outcomes. A key element to understanding how to do this is to examine the impact that visual cues – such as an age progressed picture of themselves – and a mental exercise which forces them to think critically about the future can have on their ability to engage in long-term thinking and planning.

### **24. Evaluating the Efficacy of Mixed-Mode Intercept Surveys for Complex Questionnaires**

Orin T. Puniello, *Bloustein Center for Survey Research*

This paper explores mixed-mode intercept survey data collection. The use of intercept surveys is a generally accepted substitute for probability sampling when the study population cannot be defined. Transportation researchers, in particular, frequently turn to intercept surveys to collect data on public transit riders because those populations are nearly impossible to define. In an overview study of 52 transit agencies, Schaller (2005), found that difficulty in rider population delimitation compelled the use intercept survey methods to survey riders in 96% of cases.

A 2013 transportation study of the economic benefits of the Prudential Center in Newark, NJ provides the opportunity to test the efficacy of mixed-mode intercept survey data collection. The study required a complex questionnaire that probed respondent travel and spending behavior. Given the cognitive burden of the questionnaire we theorized that providing respondents multiple-modes to complete the questionnaire would increase cooperation. Therefore, the survey protocol allowed a respondent to complete the questionnaire by paper or on the internet. We hypothesized that event participants would respond positively to an

internet option because of rapid increase in smartphones, and indeed a number of participants completed the survey at the event on their smartphones. Given this expectation the survey web survey was optimized for viewing on smartphones.

However, we are concerned that the mixed-mode data collection for this study would result in significant mode effects. The study examines data collected at three sports/entertainment events at the Prudential Center between October and December 2013 to examine and quantify mode effects. The information from this study will help inform future researchers about the efficacy and appropriateness of mixed-mode data collection in the intercept survey context. Because data collection is still underway, descriptive and analytic results are not yet available.

## **25. Consolidated Response Rates for RDD Dual-Frame Sampling**

Robert Montgomery, *NORC*

Phillip J. Smith, *Centers for Disease Control and Prevention*

Kirk Wolter, *NORC*

Meena Khare, *National Center for Health Statistics, CDC*

David Yankey, *Centers for Disease Control and Prevention*

The National Immunization Survey (NIS) shifted to a dual-frame survey design in 2011. Since that time, the survey has reported response rates separately for each sample type – landline RDD sample and cell-phone RDD sample. In this presentation we investigate alternative consolidated response rates that reflect the response outcome of the combined dual-frame sample. These alternatives include weighted combinations of the separate rates, where the weights are determined either by the relative sample sizes or by estimates of the sizes of the populations of interest. The alternatives are each based on the general principle the response rate should equal the ratio of the number of completed interviews to the number of eligible cases. We report and compare options for consolidated response rates using 2012 NIS data.

## **26. Drop-downs or Text Boxes?: Results of Usability Testing on Date of Birth Entry Method for the Online American Community Survey**

Marylisa Gareau, *U.S. Census Bureau*

Kathleen Ashenfelter, *Customer Experience Insights, State Farm Insurance*

Research regarding the use of Drop-down menu and Text-entry methods for collecting Date of Birth information on web-based surveys is mixed. Usability testing was conducted on the American Community Survey (ACS) web instrument to evaluate drop-down box method versus text-entry method for reporting the dates of birth of respondents and household members. When possible, official dates of birth for participants were gathered prior to testing from official IDs (e.g. Driver's Licenses). Participants reported their dates of birth for themselves and all household members on a pre-screening instrument. Then they completed the Date of Birth portion of the ACS for both the drop-down entry and again for the text-entry for all household members.

While the drop-down entry method limited the quantity and types of possible errors, results suggest no detectable difference in accuracy (as measured by participants' dates of birth on official ID) between the drop-down and text-entry methods. However, according to debriefing measures, the drop-down entry was preferred over text-entry method, despite negligible differences in efficiency between the two methods.

Older adults showed the same method preference as younger adults, although they were slower on average to complete text-entry dates of birth than younger adults and appeared to recover more slowly than younger adults from mistakes on the text-entry method. Error rates (as measured by mistakes, changes, and mismatches between official and reported dates of birth) for the text-entry method were much higher than for the drop-down entry method, and the highest error rates were seen for entry of non-related roommates, especially among college students. Usability findings indicated lack of clarity about the procedure when the date of birth was unknown, resulting in guessing and missing data. This poster presents the design, the methods used to detect usability error, and the findings for each of the methods.

## **27. Using Statistical Exploratory Graphical Analyses to Quickly Determine the Effect of Reducing the Number of Items in a Questionnaire Sub Scale**

Jan Beckstrand, *National Center for Organizational Development*

Boris Yanovsky, *National Center for Organizational Development*

Katerine Osatuke, *National Center for Organizational Development*

Large-scale employee surveys of workplace climate offer valuable information for informing organizational improvements (Harter, Schmidt, Hayes, 2003; Carr et al., 2003). Their challenges include assessing all relevant aspects while keeping questionnaires relatively short, to minimize the burden and encourage participation (e.g. Nagy, 2002; Rohland, Kruse, Rohrer, 2002; West, Dyrbye, Shanafelt, 2009). Standard statistical indicators that assess effects of reducing the number of items within scales are often ineffective for numerous reasons, especially when measures include restricted-range ordinal scales (e.g. 5-point, Likert-type). We present an application of Statistical Exploratory Graphical Analyses to determining the effects of dropping items from an existing scale within the Veteran's Hospital Administration's census, "All Employee Survey" (Osatuke et al., 2012). We compared composite scores and relationships among items in the AES Civility scale (Osatuke et al., 2009), reduced from eight to four items in 2013. We found that standard statistical indicators such as bivariate polychoric or other bivariate linear correlations may not adequately capture complex nonlinear relationships among items. The underlying statistical assumptions may be invalid, especially with highly skewed data. Further, even with very large groups and relatively short scales (e.g. 8-items), the sparsity of the multidimensional data matrix across the items severely limits scale-appropriate statistical analyses. Finally, such analyses require computing capacity beyond what is common in institutions. Statistical Exploratory Graphical Analyses, a relatively new method that we present, provides a viable alternative able to clarify the structure of data and discover relationships among items which can inform potential interventions. For example, in our results, respondents who rated the "fairness" item unfavorably accounted for 60 to 90% of those who unfavorably rated the other seven items, a relationship validated in the previous 4 years of data. The method allowed ascertaining no "clinically significant" changes to the previous stability and comparability of scores.

## **28. Driven to Adapt: An Application of Adaptive Design with Multiple Low-Productivity Telephone Samples**

Thomas M. Guterbock, *Center for Survey Research, University of Virginia*

James M. Ellis, *Center for Survey Research, University of Virginia*

Deborah L. Rexrode, *Center for Survey Research, University of Virginia*

Casey Eggleston, *Center for Survey Research, University of Virginia*

Darrick Hamilton, *New School of Social Research*

William A. Darity, Jr., *Duke University*

The National Asset Scorecard and Communities of Color project (NASCC) is a detailed telephone survey designed to better understand the asset and debt positions of various ethnic and racial minorities whose wealth status is often overlooked or inadequately measured. Nearly 3,000 interviews were collected in Boston, Los Angeles, Miami, Tulsa and Washington. Challenges included low incidence target populations tending to be wireless only, an intrusive survey instrument over 40 minutes long, and 85 sample quota cells. We originally conceptualized the project as a complex but traditional survey and sampling exercise. But over the course of three years and nearly 14,000 calling hours, we had to adapt our design to meet and redefine quotas, balance sample efficiency with reasonable representativeness, and maintain a sustainable rate of interview success.

The research goals dictated a disproportionate sample focused on each city's unique demographic profile (e.g., Native Americans figured prominently in Tulsa, Haitians in Miami, etc.). We used 2010 Census and ACS data at the levels of tract and ZIP Code Tabulation Area (ZCTA) to inform sampling. Lorenz curves showing geographic concentration of groups helped us identify areas for targeted sampling of some groups. We used listed landline sample by census tract, random-digit dial wireless sampling by ZCTA, and listed landline sample drawn from ethnic surname lists; some groups were targeted via several sample types.

As data collection grew to 29 separate production studies with two additional partner institutions performing data collection, we had to create new tools to extract and assemble paradata to monitor our productivity, control the budget and adapt our design.

This paper describes key events and outcomes of this unintended journey into adaptive design, including performance of wireless samples enhanced with ZIP codes, production rates, sample representativeness and design effects, and ramifications for managing expectations of project stakeholders.

## **29. Using Ancillary Data to Enhance Survey Research: The Case of Vote Validation**

Wendy Gross, *GfK*

Voter turnout is quite often over-reported in survey research. Questions about voter turnout are likely to yield positive responses, yielding a higher proportion of "voters" in survey research than population estimates (e.g., Bernstein, Chadha, and Montjoy 2001; Burden 2000; McDonald 2003). Concerned with this problem, some research has compared self-reported voter turnout with ancillary data, specifically voter rolls from government sources. However, there is mixed evidence for the accuracy of survey data vis a vis validated voter turnout. On the one hand, some scholars argue that government records are more accurate

than self-reported voter turnout (e.g., Gerber and Green 2005); on the other hand, others contend that self-reports are more accurate than validated turnout (e.g., Berent, Krosnick, and Lupia 2011).

This study examines the validity and accuracy of this ancillary data using GfK's KnowledgePanel. GfK used an outside vendor to attach turnout data from official government sources as well as measure of accuracy of the match. That is, the data identify if the panel member was found in the voter rolls and the precision of match between the information from GfK and that found in government sources. We combine this ancillary data with both profile data collected for all KnowledgePanel members and survey data collected of subsamples of panel members. This design allows us to explore descriptive questions about the accuracy and variation of the validated data. We are also able to investigate two competing hypothesis about self-reported vote turnout, specifically social desirability in affirming turnout and self-selection in completing political surveys.

Taken together, the findings of this research lead to conclusions about the value of adding this ancillary data for survey researchers as well as best practices about conducting vote validation for survey respondents.

### **30. Oversampling Minorities in the National Alcohol Survey Using the Zip Code Tabulation Area File**

Shelley N. Osborn, *ICF International*

Pedro Saavedra, *ICF International*

Naomi Freedner-Maguire, *ICF International*

Kate Karriker-Jaffe, *ICF International*

Tom Greenfield, *Public Health Institute*

The 2013 National Alcohol Survey (NAS), a study funded by the National Institutes of Health, has provided data with comparable measures of alcohol consumption, associated problems, and other social variables for more than 20 years. It is based on a dual-frame landline and cell RDD sample with a designed sample size of 5,420 adults targeting a minimum of 1,100 non-Hispanic African Americans and 1,100 Hispanics. A 95% confidence interval of plus or minus 2% was deemed desirable. This paper discusses the use of a combination of geographic targeting and screening during selection to achieve sample targets, while minimizing loss of precision due to weight variation. The first approach requires identification of areas where minorities predominate, and can result in fairly high design effects due to weighting. The second approach can be expensive when screening costs are taken into account.

The design used the Zip Code Tabulation Area (ZCTA) file from the US Census to both define strata and estimate the number of respondents by race/ethnicity. Estimates were adjusted using the distribution of race/ethnicity from a recent national survey with a similar design. ZCTA were divided into four strata: non-minority, African American, Hispanic, and Mixed. First, sample was allocated to the strata proportionally to their size. Then, one half of the respondents allocated to the non-minority stratum were distributed among the other three strata. During the respondent selection process, fifty percent of non-minority respondents in the African-American and Mixed strata, and two-thirds of non-minority respondents in the Hispanic stratum, were screened out. The cell phone sample was treated in the same way; though, strata were defined according to the ethnic composition of the county. This design met target sample sizes for sub-groups and achieved a projected

effective sample size of 3,222 cases, which is sufficiently large to achieve the desired precision.

### **31. Interviewer Gender Effects on Male Attitudes**

Richard L. Clark, *Castleton State College*

There have been several studies that have examined the relationship between race of the interviewer—perceived and real—and racial attitudes. The methodological brief that I propose focuses on the gender of the interviewer and attitudes about sexual and domestic violence. In the spring of 2013, the Castleton Polling Institute conducted a survey of men related to attitudes about domestic and sexual violence against women. Respondents were asked about what constitutes domestic violence and sexual harassment as well as their impression of the prevalence of rape, sexual assault, and violence. The survey was based on a general population random survey of men, conducted by telephone.

Because interviewers were randomly assigned number through the Institute's CATI system, the gender of the interviewer for any given respondent was, in essence, randomly assigned. While there were no real experiment conditions established at the forefront of this study, this methodological brief will examine response differences by interviewer gender to assess the degree to which, if any, interviewer gender had an impact on responses to the survey.

### **32. Venting Steam: The Effect of Write-in Responses on Response Choices for Emotionally Charged Questions**

Mark Andrews, *Ipsos Health Policy Institute*

It is very common practice in survey research to provide respondents an "other/specify" option in a question where every potential response cannot be anticipated and/or included. In many cases, this practice is an efficient way to conserve space or to record responses that were unanticipated. In questions that are emotionally charged, however, it is common for some respondents to gravitate towards an open-ended option regardless of the other response categories provided. Even in questions that allow multiple responses, some respondents will overlook appropriate response choices and provide detailed feedback in a specify box. While this respondent error can be corrected through back coding, the process is time consuming and subject to coding error.

This research will examine results from a split sample experimental internet survey, where respondent will be asked the same emotionally charged questions. One group will have an "other/specify" option imbedded in the questions. A second group will have an "other" option and will be asked to specify on the following screen. The analysis of this paper will estimate the amount of respondent error for each group and will examine the demographic characteristic of those that gravitate towards "specify" options.

### **33. Older Americans and the Changing Picture of Retirement**

Jennifer Benz, *AP-NORC Center for Public Affairs Research*

Matt Sedensky, *AP-NORC Center for Public Affairs Research*

Trevor Tompson, *AP-NORC Center for Public Affairs Research*

Becky Reimer, *AP-NORC Center for Public Affairs Research*

Emily Alvarez, *AP-NORC Center for Public Affairs Research*

Daniel Malato, *AP-NORC Center for Public Affairs Research*



In the U.S., the population aged 65 years and older is growing rapidly. In 2010, these older Americans comprised 13% of the U.S. population. Population projections for 2030 show a marked increase with older adults comprising 19% of the population, which represents an estimated 72 million older adults. Furthermore, the number of older people choosing to remain on the job has been ticking upward since the late 1990s and they now represent the fastest-growing segment in the country's workforce. By 2020, an estimated one fourth of American workers will be 55 or older, up from 19 percent in 2010. With funding from the Sloan Foundation, the AP-NORC Center conducted a national survey of 1,024 adults ages 50 and over. This survey illuminates a slow-moving shift in the American idea of retirement. Retirement is not only coming later in life, but for many it no longer represents a complete exit from the workforce. Among those older Americans who are working, 9 of 10 say they are satisfied with their jobs and more of them report positive consequences of their age at work, like having colleagues come to them for advice, more often than they report negative consequences, like being passed over for advancement because of their age. Nearly a quarter of older Americans have been on the job market in the past five years and a majority of them found the experience a difficult one. Over half of those older Americans who tried to find a job had trouble finding jobs with adequate salary or benefits and a significant minority felt too old for the available jobs, encountered employers being concerned about their age, or were told they were overqualified.

## **Thursday, May 15, 3:00 p.m. – 4:00 p.m. Demonstration Session #1**

### **Demonstration Session #1**

#### **Innovative Survey Data File Development and Production System**

Lois Timms-Ferrara, *Roper Center for Public Opinion Research*

Marc Maynard, *Roper Center for Public Opinion Research*

The process of editing survey data for release to the public is complicated and costly, and it can affect survey estimates. Based on our cumulative experience acquired over the ten-year span of the California Health Information Survey (CHIS), we have developed a generalizable system for survey data file development and dissemination. The system performs data editing, augmentation and production in a more intuitive way. We have created an innovative process that separates survey content from data processing in order to eliminate the need for intensive programming and detailed knowledge about the questionnaire. Variable content and attributes, such as skip patterns, are managed in Excel tables, and sophisticated SAS Macros read information from these tables to perform quality control checks, error detection and correction, and imputation of missing data. Excel tables are used to manage in many processes including upcoding open ending questions, generating constructed variables, as well as disseminating analytic data to ensure that respondent confidentiality is not breeched. The system is also built to incorporate quality control measures in each step of the data development process. The focus of our efforts was to standardize, streamline and modulate the process in order to create consistent, high-quality data for public consumption. We demonstrate that this system is flexible, robust, and automated and can be adopted by other surveys.

#### **3D Simulation in Survey Data Collection: Today's 3D Simulation Technology Makes it Possible to Offer Respondents the Opportunity to Receive Presentations That Are Not Only Realistic But Also Dynamically Varied to Fit Study Needs**

John Holloway, *RTI International*

Virtual reality applications are intended to simulate features of real-world environments, enabling researchers to immerse respondents into a hypothetical, real or hybrid context and study their behavioral responses to environmental cues otherwise difficult to assess in a real-life setting. Today's 3D simulation technology makes it possible to offer respondents the opportunity to receive presentations that are not only realistic but also dynamically varied to fit study needs. This presentation will demonstrate the use of a 3D web-hosted virtual environment for data collection.

This environment was created to simulate a store shopping experience. We designed a virtual store with varied configurations in which tobacco product displays and ads were present or not. We gave youth a specific shopping task and assessed whether youth exposed to the display or ad ban conditions were less likely to attempt buying tobacco products in the virtual store. The study found that banning tobacco product displays decreased youth's attempts to purchase tobacco products in the virtual store but banning ads had minimal impact on youth's purchase attempts.

Using 3D graphics development tools and presentation runtime software we are able to present scenarios that are hypothetical, real or a hybrid of the two. This broad flexibility permits the

researcher to evaluate respondent choices, behavior and impressions of presented content. This technology also permits us the opportunity to dynamically tune the presentation content based on qualifying data collected prior to the respondent's entrance into the virtual environment.

The presentation will demonstrate the virtual environment design and the utility of the 3D tool, in its application to the youth tobacco shopping simulation and discuss further survey topics that may benefit from the use of 3D interactive environments as a support mechanism including those measuring product placement, packaging, and label comparison.

## **Thursday, May 15, 4:00 p.m. – 5:30 p.m. Concurrent Session B**

### **Questionnaire Design: One Question or Many: Effects of Question Grouping**

#### **The Grouping of Items in Mobile Web Surveys**

Aigul Mavletova, *NRU Higher School of Economics*

Mick P. Couper, *University of Michigan*

There is some evidence that breaking a survey into several pages may increase breakoffs in mobile web surveys compared to a scrolling design where all items are presented on the same page. However, to our knowledge, there is almost no empirical evidence on the optimal number of items or questions to include on a page in a mobile web survey. We investigate the effect of the number of items presented on the screen on data quality in two types of questionnaires – with or without user-controlled skips. Three scrolling options were compared in a survey with 30 items where 5, 15, or all 30 questions were presented on a page. We conducted a randomized experiment using a volunteer online access panel in Russia. Panelists were invited to complete the survey via mobile phone, with random assignment to one of the six conditions. A total of 2,032 respondents completed the survey with an overall completion rate of 26.3%.

We found that displaying all 30 items on a page reduced the breakoff rate by almost a third compared to presenting 5 items per page in the questionnaire without skips. In both surveys with and without skips the completion time was significantly lower in the 30-items per page condition; however, it also resulted in higher overall item nonresponse rate. No effect was found for the number of the items displayed on the screen on measurement error in both closed-and open-ended questions. Though the questionnaire with skips did not produce a higher breakoff rate than the questionnaire with skips, there is evidence of motivated underreporting when participants could skip two open-ended questions in a row. We will present our research findings and offer some practical recommendations for questionnaire design for mobile web surveys, depending on the research goals.

#### **Question Grouping and Matrices in Web Surveys: Using Response and Auxiliary Data to Examine Question Grouping Decisions**

Ipek Bilgen, *NORC at the University of Chicago*

Michael J. Stern, *NORC at the University of Chicago*

Due to the time and cost efficiencies associated with web surveys as well as the increase in internet use among American households, survey organizations in the U.S. are increasingly using web surveys. Additionally, the increase in smartphone and tablet use makes the web surveys an attractive mode to be used in addition to the more conventional data collection modes. Hence, web questionnaire design issues, as well as device-specific visual design issues have become an important yet poorly understood aspect of survey research. One of the questionnaire design issues relevant to web surveys is question grouping and utilization of grid matrices. Studies on web surveys have illustrated that the question grouping decisions may be associated with break-offs, item nonresponse, and questionnaire completion times. In this study, we are comparing item nonresponse, questionnaire completion times, and break-offs among

three version of the question groupings and in different devices. In order to do so, we have implemented an experiment among 4,000 sampled addresses including a series of survey invitation mailings including a link to the web survey, as well as the instructions on how to complete this survey. In this experiment, three different visual groupings of several health related questions are examined: 1) item-by-item version in which each question is asked on a different screen; 2) small matrix version in which questions are presented via grids on two screens; and 3) full matrix version in which questions are presented via one grid on one screen. Our research aims to shed light on the impact of visual complexity of matrices and utilization of question grouping and grids in web surveys, as well as the comparability of different visual groupings among multiple devices with different screen sizes.

## **Achieving Balance: Understanding the Relationship between Complexity and Response Quality**

Rebecca J. Powell, *University of Nebraska-Lincoln*

Antje Kirchner, *University of Nebraska-Lincoln*

To avoid confusion with survey questions, researchers aim to create easy questions by constraining question characteristics—e.g. the Flesch-Kincaid Reading Level—in such a way to ensure that the majority of respondents will understand the question (Dillman, et. al., 2009). However, even controlling for specific characteristics, questions can still be complex. For example, two questions can receive the same reading level, but one question might be a harder task for the respondent if it asks for a behavioral frequency while the other could ask for the respondent's sex. With many surveys moving toward self-administered modes, issues of question complexity become more important since the respondent needs to comprehend all questions without the aid of an interviewer. Researchers have examined how question characteristics in self-administered surveys affect specific response quality aspects such as response times (Yan and Tourangeau, 2008). However, more research is needed to understand the extent of the relationship between complexity and response quality. This study uses the internet component of the Gallup Panel to model the relationship between question complexity and response quality. The complexity measure incorporates page level statistics, such as the number of questions, the reading level, and the type of questions. This index is then used in a cross-classified model to understand the relationship between complexity and different aspects of response quality. Since this study uses data from the Gallup Panel web survey, response quality is measured both in terms of substantive answers (e.g., nondifferentiation, “don't know” responses) and paradata (response latency, number of answer changes). Preliminary results using one wave from the internet component of the Gallup Panel show that more complex pages lead respondents to have more answer changes per page (correlation of 0.14) and have more don't know answers per page (correlation of 0.11).

## **Lean Forward: Effects of Response Format in Self-Reported Voting**

Randall K. Thomas, *GfK Custom Research, LLC*

Frances M. Barlas, *GfK Custom Research, LLC*

Self-reported voting in prior elections is often useful for examining the relationships between political attitudes and political behaviors. When asking about participating in prior elections we are often faced with a decision about the particular type of response format to use. Common formats include the Multiple Response Format (MRF or ‘select all’) and the Yes-No Grid. Prior research has shown that the Yes-No Grid typically leads to higher rates of endorsement. Since reporting behavior that may have occurred in prior years may be difficult for respondents, especially for off-year elections, we thought that an added response of ‘Not sure/Cannot

remember' might reduce the endorsement of 'Yes' in the Yes-No Grid. We conducted 2 experiments, separated by two years with respondents from an opt-in non-probability panel. Respondents were asked about their voting in the prior 2 presidential and 2 mid-term elections. They were randomly assigned to either a Yes-No Grid, a Yes-No Grid with a 'Not Sure' response, or an MRF. For both studies, we found that the both Yes-No Grids showed higher rates of endorsement than the MRF, but with some differences with the 'Not sure' response. We examine how these results would affect a likely voter model and predictions of voting behavior.

## **The "Eyes" Have It!: Evaluating Survey Instruments Through Eye-Tracking**

### **Using Eye Tracking to Examine the Visual Design of Web Surveys**

Quan Zhou, *University of Nebraska-Lincoln*

Kay Ricci, *University of Nebraska-Lincoln*

Kristen Olson, *University of Nebraska-Lincoln*

Jolene Smyth, *University of Nebraska-Lincoln*

Studies show that visual design influences how a respondent processes survey questions (Dillman et al., 2009). Sometimes the visual presentation of response options are manipulated to save space on a paper survey and to make all responses visible without scrolling on a Web survey (Couper, 2008). However, few studies have formally analyzed the effects of these sorts of graphical manipulations. In this study, we use eye tracking methods to examine the visual layout of response options in two visual design experiments: (1) a single column versus two columns of response options and (2) a linear versus non-linear five-point rating scale. Specifically, we will examine if the different layouts affect if respondents read through the response options, how long respondents spend on each subgroup of response options, and the distribution of responses selected. Data for this study will come from a general population laboratory experiment that will be conducted between November 2013 and March 2014. Preliminary analysis from a pilot study of 41 college students revealed mixed results. Initially, students spent more time looking at the one column of responses and the non-linear layout of the rating scale; on a later question in the survey, however, students spent more time looking at the two columns of responses and the linear layout of the rating scale. Although previous research has suggested that the non-linear rating scale will encourage respondents to process the options horizontally, we have observed that respondents tend to start from the middle column and use it as an anchor when processing the other options. We will investigate these trends further by analyzing these experiments with a larger sample from the general population. This study will conclude with implications for Web survey design.

### **User Experience and Eye-Tracking Study: Paper Diary Design Decisions**

Lauren Walton, *The Nielsen Company*

Jennifer Romano Bergstrom, *Fors Marsh Group*

David Hawkins, *Fors Marsh Group*

Christine Pierce, *The Nielsen Company*

Asking respondents to record their activity in a diary can be a difficult task due to retrospective reporting and cognitive burden as well as the complexity of the data collection tool. Diary questionnaires typically require multiple pieces of information including demographics, activities, and duration over a data collection period. Like other questionnaire types, visual design

principles can be used to help people perceive and understand what is being asked of them during diary measurement. Eye tracking, a technology that allows us to passively study people's eye movements, has been used mostly for questionnaire testing within the survey research field (see Redline and Lankford 2001, Graesser et al. 2006, Galesic et al. 2008, and Lenzer, Kaczmirek, and Galesic 2011). This study focuses on using eye-tracking and other usability measures to analyze how respondents perceive, understand and experience different designs of the paper Nielsen TV Diary (n=74). In this study, we use eye tracking to gain insights into visual elements that draw attention, the amount of text that respondents read (e.g., terms/instructions), and how respondent go about completing the survey. This paper centers on the collecting and analyzing qualitative and quantitative measures of the user experience, including eye-tracking data (e.g., fixation count and time to fixate), participants' verbalizations, self-reported satisfaction, and performance data (e.g., accuracy and steps to complete). The paper will also provide recommendations about the design of the paper diary based on the results of this study. Some positive findings include examples are helpful and an appreciation from participants that "their opinion matters". High priority usability issues include participants are unable to recall some information asked for in the questionnaire, and some design aspects of the questionnaire are perceived as outdated. Eye-tracking data also indicates differences in the linear gaze processing across diary design versions.

## **Eye Tracking the User Experience of a Smartphone and Web Data Collection Tool**

Kelly Bristol, *The Nielsen Company*

Jennifer Romano Bergstrom, *Fors Marsh Group*

Michael Link, *The Nielsen Company*

The visual design of survey questionnaires and other types of online data collection tools can have a powerful impact on how users perceive, understand and navigate the instrument as well as the responses they provide. Two key considerations that impact how users experience online data collection tools are the device they are using (e.g., smartphone, tablet, computer, etc.) and the method of interaction (e.g., website, app or both). When designing online data collection tools, creating a common user experience across these different devices and methods of interaction is important to help minimize any measurement differences resulting from how these features impact the responses people provide. This study draws on eye-tracking and performance data, as well as comments; from 36 participants to understand how people perceive, navigate and respond to an electronic self-report diary. We compare the user experience across four different combinations of device and method of interaction: (1) desktop PC-website, (2) smartphone-app, (3) tablet-app, and (4) tablet-website. We identify design and user experience elements that must be unified across all devices as well as elements that might need to be customized for difference devices or methods of interaction. Findings suggest that across devices and methods of interaction the visual presentation of key tasks and response options needs to be constant to ensure data accuracy. For example, some tasks in the data collection tool utilize different data entry flows for the app and website. The website layout resulted in an increased number of user errors compared to the app and comments about being confused by the task. Findings also suggest some inherent device elements, such as keyboards, may require customization. These results have implications about how to incorporate devices and methods of interaction into the design of data collection tools to minimize associated measurement differences.

## **Respondent Processing of Multiple Images throughout a Web Survey**

Nuttirudee Charoenruk, *University of Nebraska-Lincoln*

Mathew Stange, *University of Nebraska-Lincoln*

Most research on images in web surveys has examined how images affect reports to questions (i.e., visual context effects; Couper, et al. 2004). Research gaps remain, however, especially with respect to how respondents process images in surveys. Some research has examined how the placement of images in a survey affects whether respondents fixate on images and how fixation affects reports (e.g., Tourangeau, et al. 2005). What remains unanswered is how respondents process an image when it is the first, second, third, or fourth image to appear in the survey and whether subgroups of respondents process images differently (e.g., gender, age, literacy level). We answer these questions using data from an eye-tracking study that included images placed next to four questions throughout the questionnaire. We will: 1) examine if the percent of respondents who fixate on images and the amount of time spent looking at them differs by when the images appear in the survey, 2) examine whether these vary by subgroups, and 3) test if visual context effects require fixation on images and/or longer fixation or if they can occur even when respondents only peripherally process the images. This paper examines image processing and visual context effects on adult respondents from the general population with data collected between October 2013 and February 2014. Preliminary analysis from a pilot study with college students (N=41) indicates that the percentage of respondents who fixated on images declined after the second image in the survey ( $p=0.015$ ). We also found that respondents who fixated on images reported spending more money on vacation than those who did not fixate on the images, suggesting that visual context effects may differ by whether respondents fixate on images. Our paper ends with recommendations about using images in web surveys.

## **Using Eye Tracking to Evaluate Email Notifications of Surveys and Online Surveys Collecting Address Information**

Erica Olmsted-Hawala, *U.S. Census Bureau*

Elizabeth Nichols, *U.S. Census Bureau*

Using eye-tracking methodology, the US Census Bureau's usability lab is investigating how respondents interact with both email messages as well as design layouts for collecting addresses.

Email is a new way for the Census Bureau to communicate with our respondents. Both the Decennial Census and the American Community Survey are considering using email as a survey notification method. Initial usability testing of possible emails used eye-tracking techniques to understand what users read, what portions of the email they missed, what portion of the emails they spent more time on. The presentation will share eye-tracking results on the email messages.

In addition to testing possible census and ACS emails, the usability testing also included testing of address screens. For a census study, eye tracking was used to compare two different address collection designs to identify which design worked better for the respondent. We examined whether respondents saw and read the address question, the instructions, and other input fields. Results include analysis on the number of fixations per character. The eye-tracking analysis can be a good measure to identify areas of the screen that caused confusion, areas that worked well, and areas of the form that were missed and need to be re-designed.



## **Global Public Opinion Tracking as a Measure of Societal Change: Voices from a World in Motion**

### **Impact of the Arab Spring on Research in the Middle East/North Africa Region**

Anita Pugliese, *Gallup, Inc.*

Travis Owens, *Gallup, Inc.*

As the Arab Spring spread across the Middle East and North Africa (MENA), Gallup was in the field in these countries, tracking public opinion amid the unrest as it had before the uprisings. We will present our regional research trends, emphasizing overall well-being, ratings of personal safety, confidence in national government, and other measures of government performance in pre- and post-Arab Spring Tunisia and Egypt. For example, in both countries in 2012, sentiment is far from positive. Few Tunisians have confidence in their present government, and they continue to struggle financially. Egyptians are pessimistic about what their country has gained since Mubarak's overthrow.

The Arab Spring presented some unique challenges for data collection during the unrest, but it also had lasting effects on the way Gallup collects data in some MENA countries. Governmental permission to survey in some countries has become more difficult. In addition, interviewers are facing increased security challenges on the ground when conducting household surveys. For these reasons, Gallup has been transitioning from face-to-face surveys in some countries (primarily GCC countries) to phone surveys. We will describe our approach to obtaining nationally representative probability samples via landline phone and mobile phones in these countries. We will also report differences between the face-to-face and telephone modes regarding geographic coverage and respondent demographics (age, gender, education, employment status, and household size); and challenges in comparing survey responses.

### **Violence in Latin America and its Negative Effects on Trust in Institutions, Evidence from Victimization Questions from the Gallup World Poll**

Johanna Godoy, *Gallup, Inc.*

Jan Sonnenschein, *Gallup, Inc.*

Jesus Rios, *Gallup, Inc.*

Despite increased efforts by a number of Latin American governments in recent years, the region continues to suffer from some of the highest homicide rates in the world. In the late 1990s high levels of violence prompted the Pan American Health Organization to call violence in Latin America “the social pandemic of the 20th century”. In addition to the human tragedy and the loss of thousands of lives, high levels of crime and violence impose enormous social and economic costs on crime-ridden countries that hinder their institutional and economic development.

The importance of the problem of high levels of crime and violence plaguing Latin America is also reflected in a growing body of academic literature that tries to explain the phenomenon. A great part of the research is based on official homicide rates, which according to many researchers, are the only trustworthy statistics available for a longer time span. As Soares and Naritomi (2010) point out, victimization data are very irregularly distributed across countries and years. In addition, coverage in the developing world is very limited. The Gallup World Poll aims to fill this gap by offering a great wealth of victimization data from 2006 to 2013 for the most

populous countries in Latin America and the Caribbean.

Using logistic regression this paper will further contribute to the conversation about the factors associated with high levels of violence and crime in the region. Moreover, it will examine how crime undermines trust in democratic institutions and spreads fears to express political opinions. A preliminary study of World Poll data showed a strong negative relationship between levels of reported crime and trust national institutions and local police.

### **Approval of the Leadership of the Governments of Africa**

Robert Tortora, *Gallup, Inc.*

Governments across Africa take many forms. For example Angola can be thought of as a full presidential system where the president is the head of state and head of government and there is no prime minister. Other forms of government have a president as head and a prime minister that does not play an active role in government or a prime minister that plays some role as the head of government. Examples of the former include Chad and Cameroon whereas Burkina Faso, Senegal and Tunisia are examples of the latter. Kenya and Nigeria can be thought of countries with a federal form of government, a central government with an elected president that shares power with semi-autonomous regional governments. Morocco represents a constitutional monarchy where the monarch has considerable power but the prime minister is the nation's main leader. In two countries, Botswana and South Africa the president is elected by the legislature. And finally in Africa there are military junta states.

Given the wide variety of governments in Africa the purpose of this paper is to identify and compare the demographic, social and form of government characteristics of those Africans saying they approve of the leadership of their country. After a brief discussion of the issues involved in surveying in Africa data from the Gallup World Poll from 2012 and 2013 are used in two logistic regression model that use, as the independent variable, those saying they approve of the leadership of their country. Two models are developed and compared, one with the adult target population of the Gallup World Poll, those 15 and older. Since most countries in Africa have some sort of election process the second model is restricted to those of voting age, generally 18 and older.

### **Pakistan's Fledgling Democracy: Looking for Order in Chaos**

Rajesh Srinivasan, *Gallup, Inc.*

Julie Ray, *Gallup, Inc.*

President Asif Ali Zardari's election in September 2008 signaled Pakistan's complete return to civilian rule after nearly a decade under military control. Shortly after his election, many Pakistanis hoped their country would return to democracy, with a slim majority (52%) satisfied with the way that democracy was working in their country, and a plurality (28%) believing that a democratic political system was the best way to govern. But almost as many Pakistanis also didn't know what the best way to govern would be, having not experienced a stable civilian government since independence.

Although results from Gallup polls in Pakistan since then show their civilian government has never been able to engender much public trust, this discussion will present polling evidence that suggests Pakistanis' faith in other underpinnings of democracy, such as a strong judiciary, remains steadfast. At the same time, record-level support for sharia as the only source for legislation (61%) raises questions about how much the ongoing instability, in tandem with weak governance, has undermined Pakistan's course to being a full-fledged democracy.

## **Former Soviet Union: Nations, Attitudes in Transition**

Neli Esipova, *Gallup, Inc.*

Travis Owens, *Gallup, Inc.*

Countries in the former Soviet Union (FSU) have experienced a lot of change since its collapse, and many are still in transition. Even in the last two years, Gallup polls have shown respondents in 10 FSU countries are twice as likely to say that break-up of the Soviet Union harmed their country than benefited it. In Russia alone, 55% see the break-up as having harmed their country, while only 19% perceive a benefit.

These attitudes, as well as the shifts that Gallup has documented in the eight years it has been polling in the region, illustrate the challenges each country has faced as they move closer to democracy or farther away from it. In this discussion, we will examine major changes in people's perceptions about their political systems, their religiosity, what is morally acceptable, and their acceptance of people from different lifestyles and backgrounds.

Further, in several countries in the region, data collection is getting more difficult and questions that are perceived as politically or culturally sensitive may change from year to year. But in others it is getting easier, for example, since the 2012 parliamentary elections in Georgia, nearly twice as many feel free to voice their political views than they have in the past.

## **What Can Survey Paradata Do for You? Evidence from Four Federal Demographic Surveys**

### **Cost Effectiveness of Repeated Contact Attempts in the American Community Survey**

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

Deborah H. Griffin, *U.S. Census Bureau*

The American Community Survey (ACS) continuously collects demographic, social, economic and housing data to produce detailed information every year for communities across the nation. The ACS uses four sequential data collection modes (Internet, mail, phone, and personal visit), each involving multiple contacts, to raise levels of survey response. Those efforts have consistently led to high weighted survey response rates of about 97 percent. We use paradata associated with each of these modes to identify the level of effort required to obtain completed interviews. This paper focuses on the subset of sample addresses that require CAPI data collection to evaluate the cost effectiveness of additional contact attempts in this final mode, given the case's full contact history. We measure the cost effectiveness of the additional CAPI contact attempts by estimating both costs (dollars spent) and benefits (reductions in nonresponse bias). We look at these findings at various geographic levels including distinguishing among areas known to be especially hard-to-count.

## **In-Person or Telephone Collection: Mode Selection and Outcomes in the Current Population Survey**

Brian Meekins, *U.S. Bureau of Labor Statistics*

Polly Phipps, *U.S. Bureau of Labor Statistics*

Use of alternative data collection modes to optimize response and minimize cost is a well-established principle in survey design. Panel surveys are well suited to identifying modes most effective for different groups of respondents. The Current Population Survey, a rotating monthly panel household survey, is collected by personal visit and telephone. In this paper, we focus on data collection during the first four months of participation in the eight-month survey. The use of personal interview or telephone varies by month. After a household is selected, a Census field representative (FR) visits the household and collects the first month of data using computer-assisted personal interviewing (CAPI). Over the next three months, data may be collected by a Census FR via personal visit or a telephone call and entered into the electronic instrument. In addition, sample units can be transferred to a centralized telephone center that utilizes computer-assisted telephone interviewing (CATI). Eligibility for CATI is set by criteria that involve the FR, the Census regional office, and outcomes from the prior survey month. This research examines characteristics of sample units selected for CATI, verifying use of predefined criteria and identifying additional variables correlated with selection, including household demographics, survey questions, and contact and interview experience in the prior wave. In addition, we examine the success of those selected for CATI, as determined by a completed interview with a modest level of effort. Finally, we compare call center outcomes to those of units attempted exclusively in the field. Sample units not successfully completed by CATI are “recycled” back to FRs with a short time frame for collection. Information on how sample units are selected for CATI is critical to optimizing response and cost. Our analysis makes extensive use of paradata from both CAPI data collected by FR and CATI data.

## **Using Contact Histories to Predict Interview Completion in Large National Surveys**

Julia Coombs, *U.S. Census Bureau*

Rachael Walsh, *U.S. Census Bureau*

The use of paradata in adaptive design could reduce survey costs while maintaining data quality. To that end, Census Bureau interviewers have collected household contact histories on cases selected for demographic surveys for a decade. The detail of the data allows for the development of dynamic data collection procedures customized to the individual household, which could lead to the reduction of interviewer effort to complete interviews with specific households and ultimately lower field costs. Past research has predicted interview completion for simulated populations or for small samples, but the Census Bureau has contact histories for hundreds of thousands of cases from seven separate surveys. In addition, Census Bureau field interviewers often have monthly workloads containing cases from more than one survey. We include contact histories from all of these surveys and take advantage of the large sample sizes in our analysis in an effort to capture the interviewing environment accurately to better inform adaptive design procedures.

The assumptions and specifications of most models are inadequate for this type of analysis, particularly since the data are order dependent and spatially clustered. In addition, the nesting of the contact attempts within the household and again within the interviewer violates the assumptions held under traditional logistic response propensity models. Using information about the household and interviewer, we instead build a multilevel discrete time hazard model that

predicts the probability of obtaining a completed interview for the household at each successive contact attempt. By graphing resulting survival curves, we determine the optimal number of contact attempts needed to complete interviews with households. Such a model could be refined and rerun on a daily basis, leading to real-time field recommendations for interviewers and other field procedures in order to maximize interviewer effort.

## **Judgments Under Uncertainty: Evaluating Interviewer Observations of Sample Units**

James Dahlhamer, *National Center for Health Statistics*

Nancy Bates, *U.S. Census Bureau*

Rachael Walsh, *U.S. Census Bureau*

Increasingly, interviewer-administered surveys are turning to interviewer observations of neighborhoods and sample units to address nonresponse bias, whether through responsive survey design or post-survey weighting adjustments. However, quality assessments of this observational data, or paradata, have not kept pace. Since interviewer observations are largely judgments and may be incorrect, the effectiveness of responsive design strategies and nonresponse adjustments can be compromised. In this paper, we evaluate a set of new observations collected by U. S. Census Bureau interviewers with the 2013 National Health Interview Survey (NHIS). Examples of observations include assessments of the physical condition of the sample unit, evidence of smokers, and presence of children less than six years of age.

Using six months of data, we approach the analysis three ways. First, we look at interviewer compliance with completion of the observations. Are observations present for all sample cases? Second, we look at agreement within and between interviewers for a subset of cases with multiple observations (~13% of all cases). How reliable are the observations? And finally, we look at the agreement between observational and survey estimates among responding households. How accurate are the observations? For each question, bivariate analyses are followed by multi-level models that attempt to isolate the characteristics of contact attempts, sample units, and interviewers impacting discrepancies between the observational and survey measures.

Preliminary results suggest that compliance with the collection of the new observations is high, while agreement within and between interviewers for cases with multiple observations is fair to good. The level of agreement between observational and survey estimates varies and largely depends on whether the observations were recorded during a contact or noncontact. We discuss the implications of our findings for responsive design applications and nonresponse weighting adjustments.

## **Tackling Nonresponse Bias: In Search of Interviewer Observations Related to Survey Response and Survey Outcomes**

James Dahlhamer, *National Center for Health Statistics*

Renee Gindi, *National Center for Health Statistics*

Chandra Erdman, *U.S. Census Bureau*

With continued declines in survey response, federal survey organizations have turned to the collection of auxiliary measures, or paradata, to assess, reduce, and adjust for nonresponse bias. With face-to-face data collections, this includes interviewer observations of neighborhood and sample unit characteristics (e.g., graffiti, physical condition of buildings). To be effective, the

observations must be related to both the survey variables of interest and survey response. Past examinations, however, have shown this to be elusive. For example, observational measures (e.g., householder mentions of time constraints, privacy concerns) collected with the National Health Interview Survey (NHIS) have been highly correlated with survey response, but weakly correlated with key health outcomes. In this paper, we evaluate the utility of new observational measures added to the 2013 NHIS. Of particular interest are observations designed to correlate with key survey variables, including evidence of smokers, indicators that residents may be handicapped, and the age composition of the household.

Using six months of 2013 data, we start by exploring the relationship between each observational measure and survey response. We then estimate a response propensity model using promising measures identified in the bivariate analysis. Next, we focus on correlations between the observational measures predictive of response, including the response propensities from the aforementioned model, and numerous health outcomes. Finally, we undertake a preliminary nonresponse bias analysis using the response propensities from our model, grouping them into quintiles, and examining key health estimates by and cumulatively over the quintiles. Change in an estimate over response propensity quintiles is indicative of the risk of nonresponse bias, while little to no change is suggestive of the absence of bias. Findings are discussed in the context of responsive survey design and future nonresponse weighting adjustments.

## **Assessing Data Collection Effort and Nonresponse**

### **Stop Chasing Your Tail: Identifying Respondents That Are Hard to Catch and That You Will Not Miss**

Ryan Hubbard, *Westat*

Paul Guerino, *Centers for Medicare and Medicaid Services*

Increases in nonresponse and costs-per-complete in in-person interviews have forced survey managers to explore responsive designs that rely on detail case history paradata to improve cost efficiency and better allocate study resources. While responsive designs may target a number of data collection elements, there is the common goal of balancing cost and error. A number of researchers have worked to define and develop responsive design as a concept and an applied practice (Groves and Heeringa 2006, Lepkowski et al. 2010, Couper and Wagner 2011, Mitchell 2012). This paper furthers that effort through the examination of a large-scale, national survey with a rotating panel design and two decades of history spanning more than 60 rounds of data collection. The analysis focuses specifically on paradata in conjunction with demographic and frame data related to Medicare Current Beneficiary Survey late responders and initial nonresponders in an attempt to identify optimal thresholds for effort in future rounds. The results provide a cost/benefit analysis with regard to time and resources, precision of key estimates, and estimates of nonresponse bias by iteratively eliminating pools of respondents who would have been excluded had the number of attempts or conversion efforts been constrained to specific levels. Key study indicator areas include demographics, health insurance, health and functioning, access to care, health services utilization and expenditure, vulnerable populations, and satisfaction measures. This research is intended to provide information for optimizing resource allocation on the MCBS as well as a more refined blueprint for adjusting attempt and conversion efforts on other longitudinal studies to achieve a cost-error equilibrium.

## **Predicting Completed Telephone Interviews in a CAPI Environment: Can We Predict?**

Travis Pape, *U.S. Census Bureau*

Survey organizations are creating response propensity models to estimate the likelihood of obtaining a completed interview with a case. These models are often discrete hazard models, with estimates contingent on previous attempt outcomes and are being developed in both CAPI and CATI environments. This paper discusses a particular application – predicting interview completion by telephone in a CAPI survey. Interviewers are often given the option to attempt completing the interview by telephone under selected situations, such as in panel surveys. The 2013 American Housing Survey (AHS), measuring the housing stock and residential housing expenses, is a Department of Housing and Urban Development (HUD) sponsored survey conducted by CAPI of over 170,000 housing units. The ever-present cost pressures have both the sponsor and the survey managers looking for ways to reduce field costs. The lower cost of telephone interviews makes increasing the number of phone completes an attractive alternative. In this paper, we take the results of the 2013 AHS to estimate the likelihood of completing an interview by telephone on the next attempt. Covariates evaluated in the model included case characteristics derived from the sample frame, the contact attempt history and from ACS and decennial census block group demographics. We will discuss the model estimation and the practical implications for its use in data collection. We find that the model may be useful in pre-identifying cases that may be best candidates for telephone contact, and limited promise in isolating high propensity cases during data collection. We conclude with next steps toward applying the results of this study.

## **Assessing the Relationship between Number of Survey Contacts and Measurement Error using Confirmatory Factor Analysis**

Morgan Earp, *U.S. Bureau of Labor Statistics*  
Jennifer Hunter Childs, *U.S. Census Bureau*  
Rajesh Srinivasan, *Gallup, Inc.*

Is measurement error a function of response propensity? Do multiple attempt respondents contribute more measurement error than first attempt respondents? Using a measure of the public's trust of official statistics and attitudes towards use of administrative records, developed by the Census Bureau and several other key statistical agencies, we assess and compare measurement error variance between what we call early respondents (1 to 3 attempts) versus later respondents (4 or more attempts). While further attempts may help us to slightly increase our response rate and potentially reduce nonresponse bias; it may also increase measurement error. This paper assesses the impact of multiple contact attempts on measurement error variance. Measurement error variance was modeled using Confirmatory Factor Analysis.

## **AAPOR and Chapter Student Paper Award Winners Presentations**

### **2014 Seymour Sudman Student Paper Award Winner**

#### **Parenthood and Voter Turnout: Children as Resource Drain or Mobilizing Force?**

Nikki L. Graf, *University of Wisconsin-Madison*

Parenthood is a life-changing experience that alters individuals' resources, interests, and social networks, factors known to influence political participation. Despite its potential importance,

much remains unclear about the relationship between parenthood and participation in contemporary politics. In this paper, I trace the contours of this relationship and assess two competing models to explain patterns of political participation among parents. One model portrays children as draining resources such as time and money, which constrains parents' ability to participate in politics. The other focuses on how having children may shape parents' interests, expand their social networks, and motivate them to participate. Using data from the Current Population Survey (CPS) November Voting and Registration Supplement and Civic Engagement Supplement, I examine voter turnout in the 2004 and 2008 U.S. presidential elections and civic participation in 2008. Results indicate that parents of school-age children vote at higher rates than people with no children at home, whereas parents of preschool children vote at lower rates, net of other factors. Analyzing variation by gender, employment status, marital status, education, and income, I find that the negative effects of preschool children on voting are most pronounced for women employed full-time and for single parents, groups who face the greatest time strain in parenting young children. Parents' higher levels of civic participation—particularly in school/neighborhood groups—largely account for the heightened turnout associated with having school-age children. Overall, findings demonstrate that while children can have negative consequences for voter turnout in some circumstances, they can also act as a mobilizing force in parents' political lives.

**MAPOR Student Paper Award Winner – Public Opinion**  
**The Effects of Self-Interest Frames on Policy Preferences**

David Sterrett, *University of Illinois at Chicago*

Politicians and the media often highlight the possible effects of a policy on individuals, and elites craft messages that appeal to people's self-interest during policy debates. However, most of the studies on self-interest in public opinion do not examine the influence of elite messaging on people's policy preferences. Likewise, most of the framing studies in political science do not focus on messages appealing to people's self-interest. In order to address these shortcomings in the literature, this paper tests the affect of self-interest frames on people's attitudes. This paper uses data from two sets of survey framing experiments exploring people's support for proposed reforms to health care, taxes, and immigration policies. The first set of data features six framing experiment questions included in an April 2013 exit poll of 1,000 registered voters in three Chicago suburbs. Each experiment question features a control group (no frame) and three treatment groups (self-interest frame, value frame, and competing self-interest and value frame). The second set of data comes from nine framing experiments scheduled to be included in a December 2014 Web survey of 2,250 people that has already received funding and IRB approval. For each question, there will be variations in the strength of the self-interest frame (magnitude and certainty of effects) and the information environment (single message or competing messages). The results will demonstrate how self-interest frames can influence people's preferences. The findings will show whether the strength of frames and/or the information context impact the effects of self-interest frames on preferences. The experiments also explore how the influence of self-interest frames varies based on the issue and political characteristics of respondents. This paper will highlight the importance of incorporating framing into studies examining the influence of self-interest on policy preferences.

**MAPOR Student Paper Award Winner – Survey Methods**  
**Check the Phone Book: Testing Information and Communication Technology (ICT) Recall Aids for Personal Networks Surveys**

Yuli Patrick Hsieh, *Northwestern University*



Social scientists have studied personal networks for decades to understand the implications of one's interpersonal environment for behavioral and psychological well-being as well as opinion and attitude formation. Researchers in these areas mainly employ the "name generator" procedure (Burt, 1984) to collect self-reported information about who comprises survey respondents' personal networks.

However, self-reports on personal networks present a data quality challenge. The accuracy of the recalled information due to the heavy response burden imposed on respondents prevents researchers from properly assessing the size of respondents' personal networks. Thus far, although a few techniques for reducing the respondent burden have been proposed, there is little work addressing the use of memory aids that improve the recall accuracy for personal network data collection.

This study developed two recall aids for the name generator procedure and testing their utility empirically. A Web survey experiment was administered to 447 college students randomly assigned to three conditions to gather information about their personal networks. In the experimental conditions, the questionnaire encouraged respondents to consult the actual contact records stored on various communication devices (such as the "phone book" stored in a mobile phone). The results suggest that, as compared to the participants being presented the name generator in the original format, respondents who were prompted and probed to consult their devices provided more comprehensive personal network data, especially more weak ties, without substantial tradeoff of satisficing (item nonresponse and short response time) in response to the questions following the experimental manipulation. Respondents in all conditions appeared to expend a similar level of cognitive effort to report individual contacts, suggesting that recall aids were able to obtain more information without an increase in respondent burden. This study concludes with implications of ICT recall aids for research on personal networks, questionnaire design, and public opinion surveys.

### **DC-AAPOR Student Paper Award Winner**

#### **Testing for Measurement Equivalence in the Survey Responses of Incentive and Control Group Respondents**

Rebecca Medway, *American Institutes for Research*

Prepaid cash incentives consistently lead to significant increases in survey response rates. However, researchers have expressed concern that – by bringing less motivated respondents into the sample – prepaid incentives also might have unintended, negative effects on the quality of survey responses. Typically, this possibility has been investigated by comparing the responses to individual items provided by respondents who have received an incentive with the responses of those who have not received one. For groups of items that are intended to measure a latent characteristic, there is an available analytic method that may offer a better approach – testing for measurement equivalence. This method has been applied in other areas of survey methods (such as mode or cultural comparisons) but has not yet been applied to incentives research. This paper extends the existing incentives literature by conducting a measurement equivalence analysis for three multi-item scales drawn from three recent surveys that included prepaid incentive experiments. Overall, the results are reassuring for practitioners, as they suggest that measurement equivalence largely was established between the responses of incentive and control group respondents.

## **Administrative Record Linkage in the United States and Europe: Methods to Mitigate Bias and the Merits of Linked Data**

### **Who Consents to Record Linkage: Can Incentives Mitigate Bias?**

HarmoniJoie Noel, *American Institutes for Research*

With response rates declining over time (Groves, 2011), researchers are turning to administrative records as a supplemental method for collecting rich and comprehensive data from study participants. However, such linkage requests typically require obtaining consent from study participants, and are therefore susceptible to nonresponse bias. This paper reports the findings from a mail-based study evaluating the impact of different levels of incentives on obtaining participants' authorization to release their Social Security Administration employment and benefits records for survey research. This paper specifically examines the potential nonresponse bias related to record linkage consent and evaluates the extent to which incentives alleviate this bias.

Other record linkage studies have shown differences between consenters and nonconsenters on demographic characteristics as well as study outcome measures (Sakshaug et al., 2012). Preliminary analyses from this study show that among survey completers, consent rates are higher for those who are more educated, reported higher levels of civic engagement, were more cooperative, and had more interest or familiarity with the study (Stone et al., 2013). We also found that the \$20 incentive increased consent rates over the \$0 and \$2 incentives. This paper builds upon the previous analyses by examining the interactions between respondent characteristics associated with consent propensity and incentives to answer the question: "Do financial incentives help to offset the bias found in record linkage consent, in particular the pro-education bias?" Findings from this paper could help promote the use of linked data and inform researchers' use of incentives in studies that require consent for record linkage.

### **Propensity to Consent to Data Linkage: Experimental Evidence from the Innovation Panel on the Role of Three Survey Design Features**

Jonathan Burton, *University of Essex*

There is an increasing amount of methodological research on eliciting consent to link administrative data to survey responses in surveys. However, most of the survey design decisions regarding asking for consent are often based on anecdotal accounts and practical issues relating to the implementation of the survey, rather than sound empirical evidence.

The aims of the paper are threefold: it aims to: (i) investigate the reasons why people consent/do not consent to data linkage, (ii) explore the effects of the position of the consent question in the questionnaire on response rates and bias, and (iii) confront and contrast the effects of different ways of re-asking consent (i.e., using dependent and independent questions) on consent rates and bias. We also present initial results from an analysis of the interviewer-respondent interaction, elicited from audio recordings of the interview. We report findings from wave 4 of the Innovation Panel (IP) of Understanding Society: The UK Household Longitudinal Study.

### **Using Gain-Loss Framing to Ask Respondents for Consent to Link Survey and Administrative Data**

Joseph Sakshaug, *Institute for Employment Research*

Many survey agencies link sample survey records to corresponding records collected from administrative sources. Prior to linkage, it is important to obtain consent from respondents. Obtaining consent, a mandatory procedure in most countries, is needed to ensure respondents are informed of how their data will be used and protected. However, empirical research has shown variation in consent rates across studies and systematic differences between those who consent and those who don't. These effects can lead to possible biases in inferences obtained from linked survey and administrative data. In efforts to increase linkage consent rates in surveys, researchers have conducted experiments with various wordings of the consent request. For example, a common approach has been to emphasize the positive aspects of linkage (e.g., improved data accuracy, reduced burden) during the consent request. However, this "gain framing" approach has not led to significant effects in actual implementations. Contrary to conventional wisdom, a more effective strategy may be to emphasize the negative aspects of not providing consent. This "loss framing" approach is supported by Kahneman and Tversky's "prospect theory" which asserts that people are risk seeking when it comes to sure losses (as opposed to sure gains) and are more sensitive to losses than to gains of the same objective magnitude. We tested the effectiveness of this loss framing strategy in a telephone experiment in which half of respondents received a gain framing request while the other half received a loss framing request for linkage consent. In line with prospect theory, we find that respondents in the loss framing group agreed to the consent request at a higher rate than those in the gain framing group.

### **Interviewers' Influence on Bias in Reported Income**

Manfred Antoni, *Institute for Employment Research*

Questions on sensitive topics like income often produce relatively high rates of item nonresponse or measurement error. In this context several analyses have been done on item nonresponse, but little is known about misreporting. One possible explanation of such misreporting is social desirability bias, which may lead to over reporting of desirable attributes or underreporting of undesirable ones. Because a competent interviewer may be able to inhibit such behavior, interviewer characteristics as well as the interaction between interviewer and respondent characteristics should be important factors. We therefore examine the influence of respondent and interviewer characteristics on the accuracy of reported income.

Using linked survey and administrative data we are able to detect the extent of the social desirability bias in reported incomes and explore the influence of respondent and interviewer characteristics on it. The starting point for the linkage is data from the German National Educational Panel Study (NEPS): Starting Cohort 6, which surveys adults from birth cohorts 1944 to 1986. In addition to the survey data, NEPS provides detailed paradata, including interviewer characteristics and context data about the interview. More than 90% of the respondents consented to a linkage of their survey information with administrative data from the German Federal Employment Agency. These longitudinal earnings data are highly reliable as they are based on mandatory notifications of employers to the social security system. Both data sources were combined by record linkage techniques using the respondents' names, gender, birth date and address as identifiers.

As we include interviewer as well as respondent characteristics into our model we employ a multilevel strategy to estimate their respective impact on the incidence and size of any bias in reported incomes. This allows us to control for latent interviewer traits that might have influenced the respondent's answering behavior during each interview of a given interviewer.

## **Using Administrative Records to Develop and Evaluate Measurement Error Associated with New Survey Items in Federal Collections**

Celeste Stone, *American Institutes for Research*

Administrative records are not only useful for the data that they provide, they are also useful tools for developing and evaluating the measurement properties of new survey constructs and items. In 2010, a federal Interagency Working Group designed a pilot study evaluating new survey measures that could be used to estimate educational credentials not adequately or consistently measured in the past, despite their potential added value to education and labor market research.

The goal of the pilot was to develop and evaluate a parsimonious set of survey items that could be included in multiple federal surveys, such as the National Household Education Survey and the Survey of Income and Program Participation. Therefore, it was imperative that the study design allow for a careful evaluation of the measurement properties associated with these new measures of non-degree credentials (e.g., industry-recognized certifications and state and local government-issued licenses). The development effort culminated in the Adult Training and Education Survey Pilot Study—a household survey of non-institutionalized adults conducted by mail and telephone by using an address-based, nationally-representative sample and a convenience, or “seeded,” sample of certification or license holders who were identified using administrative records to develop and quantitatively evaluate new survey measures.

This paper describes the specific methods used, details how the methods were operationalized, presents the results, and describes how the results were used to refine the survey items and questionnaires. The study specifically examines the use of administrative records to (1) detect systematic underreporting by respondent and credential characteristics, and (2) identify credential commonalities—or groups of survey items that could be used in combination to identify instances of over reporting—that can be used in surveys where administrative records are not readily available.

## **Pre-Election Polling Methods**

### **In Search of More Granular Likely-Voter Models for Low-Turnout Elections: A Look at a Sample of 2013 and 2014 Primary Elections**

Julia Clark, *Ipsos Public Affairs*

Neale El-Dash, *Ipsos Public Affairs*

Clifford Young, *Ipsos Public Affairs*

Most public polls use some derivation of the old “Gallup” Likely-Voter model which typically includes 5 or 6 items summated into an index. Likely voters (LV), then, are defined by a “cut point” which typically corresponds to the historical turnout rate in that given election. Because of the coarse nature of the index, there are two potential problems:

- 1 – It may be impossible to obtain a LV-cut that approximates the expected turn-out.
- 2 – The predicted turn-out in two consecutive LV-cut points can be very different, not allowing the researcher to examine what happens in between these cut-points.

These problems are especially acute in low turnout elections, such as primaries. Indeed, the top 25% of declared likely voters tend to be clumped together in the top box of the scale. In a low

turnout election where only about 15% of the electorate vote, this inability to discriminate is a serious handicap. The second problem is that a 25% turnout has a decidedly different partisan makeup than a 15% one.

With these challenges in mind, we have employed estimated probabilities of voting using logistic regression as a function of past behavior, intended future behavior, and degree of partisanship. Our model provides two advantages. First, we are able to discriminate voters in one-percent intervals from 0 to 100%. Second, by employing political variables, our model captures the partisan nature of primary elections.

Our paper will examine the relative performance of the traditional summated index likely voter approaches with our logistic regression method. We will analyze approximately 15,000 interviews collected for the Reuters/Ipsos 2012 primary polls in OH and FL as well as another approximately 18,000 interviews to be collected in six 2014 primaries. To measure performance, we will employ the Average Absolute Difference between the survey estimates and election results.

## **Model-Based Survey Inference for Political Polling**

Masahiko Aida, *Civis Analytics*

Since the publication of the seminal paper by Neyman in 1934, design-based survey inference became the de facto method of survey research in most of the 20th century. However, the advantages that previously made design-based survey inference an appealing choice are quickly disappearing, due to factors such as the growth of the cell-phone-only population, declining response rates, sky rocketing costs and the relative inefficiency of the random digit dialing method (for political polling at least).

The author argues in this paper that model-based survey inference (Carl-Erik Särndal, 1978), which was in heavy use in the 2012 Obama campaign, has the potential to mitigate many of the mentioned issues that design-based surveys face today.

The model-based inference has several advantages: (1) it suffers less from nonresponse/non-coverage, provided that appropriate covariates are chosen to balance the sample; (2) it can provide far more precise estimates for small areas and subgroups; and (3) it can provide a meaningful alternative to the current likely voter screen, which is known to be quite inaccurate (Rogers & Aida, 2013).

At the same time, the author discusses potential pitfalls of using model-based inference by the means of regression estimators, as well as practical methods of testing robustness and validation of model specification.

The author predicts that the model-based survey inference will be the dominant approach in the future of political polling, and will demonstrate the example using 2012 data.

## **Pre-Election Polling in Afghanistan**

**Gary Langer, *Langer Research Associates***

Matthew Warshaw, *ACSOR*

Greg Holyk, *Langer Research Associates*

John Richardson, *D3 Systems*

Afghanistan stands on the edge of a new future. Its so-called “transformation decade,” a transition to greater self-dependence, begins with the country’s election of a new president this spring. We’ll report at AAPOR on a challenging survey research project: National pre-election polling in Afghanistan conducted in the months leading up to that scheduled election. Funded in part by the U.S. Department of State in a cooperative agreement between the U.S. Embassy in Kabul and ACSOR Surveys, our random-sample, face-to-face national surveys will explore Afghans’ living conditions; political, social and economic concerns; democratic values; and assessments of the candidates running to succeed Hamid Karzai next April 5. As with thoughtful pre-election polling in the United States and elsewhere, the aim of the project is to enrich public understanding of the concerns and choices facing ordinary Afghans in their upcoming national election and beyond. In addition to presenting research results, we’ll discuss methodological and operational challenges in conducting public opinion research in a conflict zone.

## **Voting - A Memorable Decision?**

Richard Ohrvall, *Statistics Sweden*

Mikaela Jarnbert, *Statistics Sweden*

Election forecasts and estimates of hypothetical elections are frequently produced by various political opinions polls. In many of those polls, voting and party choice in a previous election are used as auxiliary data in the estimation. A central question is whether this type of auxiliary information is of sufficient quality. Is it likely that a respondent correctly remembers voting decisions made previously in time, often many years ago? Also, some theories suggest that a respondent's memory is influenced by current party preference. This can occur due many different mechanisms, and lead to measurement error and thereby biased estimates.

Using dataset from the Swedish Party Preference Survey, we set out to study this type of measurement error. First, we analyze survey answers to questions about voter turnout in previous election. In this analysis we use validated voting, based on information from electoral rolls. Second, we try to assess measurement error in questions regarding party choice in previous election. This is facilitated by the panel design of the Party Preference survey, which allows us to study both the consistency in memory over three waves of the survey. It also gives us an opportunity to study whether or not a change of memory is associated with a change in the current party preference.

**Friday, May 16**  
**8:00 a.m. – 9:30 p.m.**  
**Concurrent Session C**

## **Public Opinion and the Affordable Care Act**

### **Tracking Public Opinion on the Affordable Care Act: From Passage to Early Rollout of Coverage Expansions**

Liz Hamel, *Kaiser Family Foundation*

Bianca DiJulio, *Kaiser Family Foundation*

Jamie Firth, *Kaiser Family Foundation*

Claudia Deane, *Kaiser Family Foundation*

The Affordable Care Act (ACA) was signed into law in 2010, but the major provisions expanding health insurance coverage, along with the “individual mandate” requiring most Americans to have health insurance or pay a fine, take effect in January 2014. Kaiser Family Foundation (KFF) tracking polls, conducted monthly since the law was passed, have found public opinion on the ACA to be remarkably stable over time, characterized by deep partisan divisions reflecting the strong pro- and anti-Obamacare messages voiced by Democratic and Republican lawmakers in Washington and around the country. Confusion and misperceptions have also been common, with many Americans unsure about the status of the law, unaware which provisions are and are not contained in the ACA, and feeling uninformed about how it will impact their own families. As the country moves into the next phase of ACA implementation, KFF will continue to track the public's views and experiences with the law through monthly tracking polls. This presentation will explore whether the stability of opinion continues into this next phase, or whether shifts occur as more people learn about the law through their own experiences or those of their family and friends. We will also explore any changes in partisans' perceptions and how those correspond to political developments related to the ACA. Finally, individual states will experience the rollout somewhat differently based on lawmakers' decisions about whether to expand Medicaid and whether the state operates its own health insurance marketplace or joins the federal marketplace. We will examine public opinion at the state level (aggregated by groups of states making similar decisions) to assess how these state-level decisions may or may not interact with the public's views of the law.

### **The Affordable Care Act: When Political and Product Advertising Collide**

Elizabeth Wilner, *Kantar Media*

President Obama's healthcare reform law is unprecedented in the annals of US television advertising for the following reasons:

As a 4-year-old government program that is the focus of more political ad spending than any other program in history except for Social Security and Medicare, which have been the subject of political ads for 10 times as long. For Republicans, eradicating or weakening “Obamacare” has basically replaced the formerly standard call for cutting Social Security and Medicare in their ad messaging. The depth this anti-ACA advertising has reached on the ballot—in races for President and US Senate on down to races for state and local offices—is what has caused



spending on ACA-related political advertising to skyrocket so fast.

As a law that broke the mold of traditional issue advertising by inspiring far more—and indeed, still ongoing—issue advertising after its enactment than before. This is due both to its implementation timetable and to its standing in public opinion. Critics of the law continue to see it as vulnerable in and of itself, and as an issue they can use to defeat Democrats.

And, as a law that directly unleashes hundreds of millions of dollars—and in the coming years, billions of dollars—in product advertising because it instantly creates millions of newly eligible customers and sends additional millions of formerly insured customers back into the marketplace. Health insurer TV ads are now airing in the same markets as political ads excoriating the law. Other product advertisers, such as tax preparation services, also are acknowledging the law.

This presentation will look at the ACA's place in political advertising history and how the messaging, spending and placement of ACA-related TV advertising, as well as news coverage and other forms of communication about the ACA, correspond over time to the evolution of public opinion about the law.

### **The ACA: Americans React as the Law Moves from Policy Debate to Reality**

Bill McInturff, *Public Opinion Strategies*

The Affordable Care Act (ACA) is moving from an abstract policy debate to the “health care reality outcome” portion of the debate where Americans’ attitudes will be determined based on their assessment of the new health care law and its consequences to them personally. The NBC/WSJ and CNBC surveys carefully built on the core tracking data available through the Kaiser Family Foundation monthly tracking to establish a detailed set of benchmark questions in September before the exchanges opened on October 1, 2013 that could measure this personal impact over time. Specifically, across these two survey platforms, we will be tracking whether people perceive their premiums have gone up, their working hours reduced, whether they have lost their health insurance coverage or were able to get private coverage for the first time as a consequence of the new health care law. As well, we will be tracking broader perceptual measures about whether the new law is having a positive or negative impact (or not much of an impact either way) on the country’s health care system or on them and their family. This data will be placed in a broader political perspective by whether the respondent lives in a congressional seat held by a Republican or Democratic Member of Congress. Finally, given our firm’s extensive political research, we will attempt to provide some outlook on how this issue might impact the 2014 elections.

### **Messaging to the Uninsured about the Affordable Care Act**

Tresa Udem, *PerryUndem*

In October 2012, one year prior to the implementation of the ACA, 78% of uninsured adults lacked awareness about new options under the law. That number remained essentially unchanged leading up to the launch of open enrollment on October 1, 2013. During much of 2012 and 2013, the federal government and state-based marketplaces conducted extensive public opinion research to identify the strongest motivators for enrolling in coverage, the top facts people needed and wanted to know, and how to talk about concepts like subsidies. Some of the research informed advertising and outreach campaigns that launched at various times during the summer and fall of 2013. Some campaigns adhered strictly to the research and some did not - or did not conduct much research. Despite public education efforts, the uninsured

remained under informed and misinformed during the first open enrollment period. PerryUndem conducted messaging research for Enroll America, Planned Parenthood, the Robert Wood Johnson Foundation, and four state-based marketplaces. This presentation will go over the major research findings around messaging to uninsured individuals to encourage enrollment. It will include key communications challenges, such as talking about the mandate and affordability of new plans. It will also address how research did or did not inform some of the state-based enrollment efforts.

## **Tracking Changes in Public Opinion of the Affordable Care Act**

Katherine Carman, *RAND*

This paper reports the results of a new public opinion tracking poll using the American Life Panel (ALP), focusing on how opinion changes over time. Our survey differs from others in three main ways. First, the same survey respondents are contacted each month. Thus we can observe whether individuals change their opinions over time, rather than just observe changes in the overall popularity of the ACA. Using longitudinal data to track public opinion helps to provide a much clearer picture of whether changes are due to errors introduced by a changing sample as used in a repeated cross section or whether actual changes are occurring. Second, the survey is continuously in the field, with a different subset of the overall sample being surveyed each week. This allows us to collect and track real time data and observe how quickly opinion changes in response to events. This means that at any time some respondents are being surveyed so we can measure public opinion of the ACA on a weekly, not just monthly, basis. Third, we survey over 5000 individuals and have a variety of background characteristics previously measured in the ALP, allowing us to relate opinion and change s in opinion to individual characteristics or characteristics of the state where they live.

We investigate the stability of individuals' opinions focusing on the impact of current events, individuals' characteristics and experiences, and state characteristics. For example, we investigate whether gaining access to insurance after being uninsured is associated with changing opinions. Additionally, we consider differential evolution of opinion based on state policies such as use of the federal marketplace website vs state websites or the decision to expand Medicare. Understanding what drives changes in opinion at an individual level will help to better understand the results from cross sectional public opinion polls.

## **Public Opinion, Partisanship and Democratic Representation**

### **Crashing the Grand Old Party's Traditional Stances on Social Issues: The Effect of Issue Framing on Public Opinion in the Context of Conflicting Value Cues**

Ashley A. Coning, *Rutgers University*

Traditionally, the Republican Party has taken hardline conservative stances on a variety of social issues. Some Republican elites and organized interests have bucked their party's positions, however, to advocate for issues that seemingly go against their partisan and ideological identities, claiming that these issues embody the very values their own party holds dear. The "value frames" these Republicans use to attempt to sway opinion among fellow partisans in an atypical direction are distinct from those used by Democrats on the same issues and from counter frames offered by traditional Republican opposition. But existing research on

value frames and their effect on public opinion has thus far only focused either on consensus values that appeal to everyone or on partisan values supporting established partisan positions.

This paper develops a new approach to framing by identifying what I call “conflicting cue” value frames, in which long-held partisan values are attached to atypical party stances. I analyze, through survey experiments, a set of conflicting cue arguments currently being used by Republican and conservative interest groups and elites in support of same-sex marriage, abortion, and immigration and assess their impact on public opinion both across and within partisanship. These “conflicting cue” value frames may have the ability to move public opinion within groups where opinion was once thought of as immovable – showing the great importance values can play in influencing issue positions.

Building upon Chong and Druckman’s competitive elite framing and counter framing theories, this research strives to untangle a new kind of framing in real world politics – especially as the GOP stands at the brink of reassessment and potential “civil war” over social change. This work has important theoretical implications for the relationship between elite communication and mass opinion formation, as well as the stability of attitudes, partisanship, ideology, and values.

## **Forecast LA: What Public and Leadership Outlook Surveys Can Tell Us About the Region**

Brianne Gilbert, *Loyola Marymount University - Center for the Study of LA*

Fernando Guerra, *Loyola Marymount University -Center for the Study of LA*

Brittany Machado, *Loyola Marymount University -Center for the Study of LA*

As cities emerge from the difficulties of the national recession, understanding the economic choices leaders will take in the coming years is essential. As mediators between government institutions and its residents, mayors must both work within the city’s strictures and organize the economic and social life of the city. Mayors must mediate issues such as race relations, property values, the matriculation and retention of businesses, access to consumer goods, and quality of transportation within the city.

Researchers at the Center for the Study of Los Angeles at Loyola Marymount University add projections on economic trends to the context of attitudes and expectations of both government leaders and members of the public whose everyday lives are affected by the state of the economy and current political climate. Using public opinion and leadership surveys they leverage what the public thinks about its quality of life and the effects of the economy and political climate with the insights leaders have on their city and its trajectory.

The Leadership Outlook Survey asks mayors of Los Angeles County’s 88 cities for their thoughts and opinions on what lies ahead for their own municipalities in the context of local, state, and national economic and political landscapes. As a baseline survey for its new annual study, the Public Outlook Survey queries LA County residents about their personal finances, quality-of-life concerns, expectations for the local economy, and other civic and political issues. These two opinion surveys will be analyzed and compared to produce insightful context for the most important issues and challenges on the region’s horizon.

By making these results available to leaders and the public, the Center will help the community understand the similarities and differences between the views of leaders and residents to help encourage civic dialogue, promote trust and cooperation, and stimulate thoughtful collective action.

## **The Importance of the Liberal-Conservative Measure when Correlations with Partisanship are Strong**

Natalie Jackson, *Huffington Post/Pollster.com*

Is ideology different from partisanship in an era of polarization in which partisanship offers clear preferences with less overlap than in the past? Theoretically, yes: ideology is not defined by political institutions, as partisanship is, and as such can explain variance within and outside of partisanship. Empirically, ideology is operationalized as a liberal-conservative measure, and its strong correlation with partisanship brings up the question of whether both measures are necessary. Using data from an original survey experiment exploring the dimensionality of the liberal-conservative measure, conducted by Knowledge Networks (now GfK) through the Time-sharing Experiments for the Social Sciences grant program, this paper demonstrates how the measure is a relatively simple heuristic which can identify nuance in opinions and preferences that partisanship cannot. Liberal-conservative scales demonstrate where, spatially, within a party an individual is located, and it can be used to differentiate types of views--e.g., explain how individuals can have distinct views on types of issues, such as economic and social, but still identify with a single party. These measures of ideology can provide a more nuanced understanding of opinions and preferences than partisanship even when correlations are strong.

## **A New Method for Gauging Public Opinion in the States: Aggregation of National RDD Surveys**

Bo McInnis, *Stanford University*

Jon A. Krosnick, *Stanford University*

The U.S. federal government is composed partly of officials who are elected to represent people living in limited geographic areas (states, congressional districts), and these officials might like to know what their constituents want government to do on specific policy issues. However, the field of public opinion research rarely provides to these officials data of direct use to them. Instead, the field tends to generate in-depth studies of public opinion in national surveys of sizes unsuitable for breakdown into states or congressional districts. This project explored a new method to generate state estimates: combining data from a series of national RDD surveys conducted using Stanford University question wordings that measured opinion on a single issue in depth. Statistical modeling of mode effects, effects of season of interviewing, effects of sponsor and data collection firm, and changes in opinions of over years allowed for the generation of estimates for almost all U.S. states using a variety of measures of opinions relevant to the issue of global warming. Statistical analyses were conducted using Pew and Gallup surveys on the same topic to test the validity of the state measurements made with the Stanford question wordings.

## **Methodological Briefs: Survey Methods in Multinational, Multicultural and Multiregional Contexts**

### **Who Is to Blame for Bribing, the Government or People?**

Data Tsabutashvili, *Gallup, Inc.*

Neli Esipova, *Gallup, Inc.*

Corruption and bribing is considered as one of the main obstacles to a country's economic development. As commonly considered, fighting corruption is difficult due to the fact that in many countries it has become the way of everyday life. Bribing at different levels of government is seen as an easy way to solve ones' personal or business related problems. We challenge the generally perceived idea that the corruption is viewed as being negative by people, with government being solely responsible for it. This study conducted in twelve Former Soviet Union (FSU) countries will focus on understanding whether long term bribing has led to the perception of bribing as being morally acceptable.

Within the scope of 2006 and 2013 Gallup World Poll in FSU countries we asked several questions about whether bribing for different reasons is morally acceptable or not. The study is based on nationally representative data with at least 1000 interviews in each country. Through Latent Class Analysis we aim to identify and estimate the sizes of latent groups with different perceptions about the acceptability of bribing. We will focus on the analysis of data from Russia but will also compare latent groups across all twelve countries.

We don't expect to find the same latent class structure in every county but believe that in all countries we will be able to distinguish the groups similar to the following: 1) a group for whom bribing for any reason is acceptable, 2) a group for whom bribing as a necessity to achieve basic personal goals, 3) a group for whom bribing is a good way to do business and 4) a group who thinks that bribing is unacceptable in any situation. We will also examine if the structure or the sizes of latent groups has changed from 2006 to 2013.

### **Measuring Attitudes toward Immigration across Two Decades: Derivation of an Acceptance Score and Trend Analysis**

Adriana Espinosa, *The City College of New York*

Glen Milstein, *The City College of New York*

The public's attitude toward immigration is a foundation of public policy development. The challenge for researchers is to calculate a national average for this controversial subject from the public's multiple attitudes. For nineteen of the past twenty years at least one of several national opinion surveys in the United States asked respondents to state their preferences toward levels of immigration. In all these surveys, participants were asked to choose one response from a hierarchy of three (e.g. 1 - Decreased; 2 - Same; 3 - Increased) or five options. By virtue of this construct, answers were analyzed separately: the annual percentage of persons designating "Decreased", those designating "Same", were separately described and analyzed as was "Increased". Three to five distinct sub-group analyses impedes realizing a unified measure of the country. In this paper we derive an aggregated acceptance score, allowing researchers to capture the dynamic changes across responses, and ascertain the spectrum of changes in attitudes towards immigration across time, using multiple surveys at once. We apply this new acceptance score to an analysis of responses from ANES, GSS and Gallup opinion polls. Our analysis demonstrates how the separate measurement categories from different national surveys can be aggregated to gauge trends in national attitudes over time. Our results

confirm other longitudinal research, demonstrating that across the last two decades the United States has become more accepting of immigrants.

## **Identifying Unauthorized Immigrants in Pew Research Center's National Survey of Latinos**

Mark Hugo Lopez, *Pew Research Center*

Ana Gonzalez-Barrera, *Pew Research Center*

Jeffrey S. Passel, *Pew Research Center*

An estimated 11.7 million unauthorized immigrants resided in the U.S. in 2012, according to the most recent estimates of the Pew Research Center. Surveying this group and learning about their views and attitudes is important for researchers and policymakers. However, surveying this population is difficult. To identify unauthorized immigrants, some surveys take a direct approach by asking immigrants about their legal status, which carries the risk of eliciting dishonest answers. As an alternative, since 2007 the Pew Research's National Survey of Latinos (NSL) has indirectly asked Latino immigrants about their legal status. This is done with two questions regarding U.S. citizenship status and Legal Permanent Residency (LPR) status.

According to prior analysis of Pew Research's estimates of the unauthorized population based on U.S. government data, above 90% of Latino immigrants who are neither U.S. citizens nor LPRs are unauthorized immigrants. Thus, we know that NSL respondents who are not-citizens and not-LPR Hispanics are very likely unauthorized.

But how well does our indirect survey method do in identifying unauthorized Hispanic immigrants? This paper evaluates Pew Research's indirect method using the 2012 NSL. First, we compare the share of NSL respondents that is likely unauthorized to the share of the Hispanic adult population estimated to be unauthorized. We find that the share in the NSL identified as likely unauthorized is considerably lower than the estimates. We then compare demographic characteristics of likely unauthorized immigrants in the NSL with those of Pew Research's estimates of the unauthorized. This analysis shows that in terms of demographic characteristics the NSL sample of likely unauthorized immigrants is broadly similar to the unauthorized population estimates. However, the NSL sample may not be fully representative as it is more female, has lower educational attainment and is less likely to be employed than the estimates of the U.S. unauthorized population.

## **Non-English Modes of Communicating Political Information: A Proposed Model on the Role of Language in Social Cognition**

Alejandro Flores, *The University of Chicago*

In the U.S., non-English modes of communication have been used for decades. Traditionally, when political information is presented in other languages, the assumed logic in their use is that they are a general way to appeal to a specific demographic. These linguistic appeals have an implied persuasive effect on social cognition, capable of influencing how an individual processes information. For example, President Obama aired television ads on Spanish networks for both his 2008 and 2012 campaigns. The popular and cursory assumption is that these non-English strategies resonated with Latino voters. If these linguistic modes of communication are designed to appeal to particular voters, what does it mean for language to work? If President Obama's use of Spanish-language ads won over Latino voters, did they work in the same way as Spanish-language ads used by President Bush? Although the role of language has been explored in the context of consumer research (claim and brand recognition), no studies have

been conducted exploring how bilinguals process political information. Individuals, however, evaluate political information very differently than merely deciding between brands. Moreover, existing studies on the role of language generally test a single language. In this light, the following question can be asked: for bilinguals, what is the role of language when communicating political information? This study utilizes experimental data with bilingual subjects, primed with non-English appeals relating to different political issues as the treatment to test a model for language. This experiment was designed to explore social cognition among different sets of bilinguals (Spanish, Hindi, and Chinese) in order to determine if the persuasive effect of these linguistic appeals operate in the same way. If individuals have a positive response to a mode of communication that “speaks” their language, then is it the content, the language, or an interaction of both the stimuli?

### **Conducting a Longitudinal Survey with Drug Treatment Patients in Afghanistan: Methodological and Contextual Challenges**

Matthew Courser, *AP-NORC Center for Public Affairs Research*

Melissa Abadi, *Pacific Institute for Research and Evaluation-Louisville Center*

Stephen Shamblen, *Pacific Institute for Research and Evaluation*

Linda Young, *Pacific Institute for Research and Evaluation*

Knowlton Johnson, *Pacific Institute for Research and Evaluation*

Kirsten Thompson, *Pacific Institute for Research and Evaluation-Louisville Center*

Amanda Bajkowski, *D3 Systems, Inc.*

Afghanistan is the world’s principal source for heroin and opium. Current estimates suggest that roughly 8% of Afghans between 15 and 64 are drug abusers (about twice the global average). In 2005 the U.S. Department of State Bureau of International Narcotics and Law Enforcement Affairs began supporting the operation of a number of residential drug abuse treatment (DAT) centers in Afghanistan. Seven of these DAT centers participated in an evaluation conducted between 2009 and 2011. The evaluation assessed pre-post changes in patient drug and alcohol use and antisocial behavior after completing inpatient treatment and the first year of outpatient treatment. The evaluation included two waves of face-to-face survey data collection and drug testing (at entry and 12 months after entry) with 504 male and female patients. Patient progress through treatment was tracked using treatment logs completed by Center staff for each patient.

Although the evaluation achieved a 70% study retention rate and found that the DAT centers were successful in reducing patient drug use and antisocial behavior, this success required the research team to address significant methodological and cultural challenges. This presentation will review key challenges and how they were resolved, along with implications and best practices for future survey work in international contexts. Challenges to be discussed include: cultural differences in the meaning of standard survey constructs (such as the significant variation in ages reported at pre- and post-test surveys because many Afghans are unaware of their age), completing interviews and monitoring fieldwork in remote conditions within an active war zone, tracking highly mobile/nomadic patients between the pre- and post-test interviews, and conducting interviews with female patients (a particular challenge during the post-test patient survey when in many cases the only private environment for the interview was at the DAT Center, requiring significant travel with a male escort).

### **Mixed-Method Assessment of Validity and Cross-Subgroup Comparability**

Kristen Miller, *National Center for Health Statistics*

Mitchell Loeb, *National Center for Health Statistics*

Over the past decade, the ability to assess cross-cultural comparability through survey design has generated increasing attention. At issue are whether questions have the same meaning to all cultural and economic groups represented in a survey and whether the data elicited from the questions capture the same phenomena across all socio-cultural groups surveyed. Small, qualitative studies have shown that respondents from different backgrounds can interpret and process questions differently in certain circumstances. For example, cognitive interviewing studies have shown that lower socio-economic respondents, particularly with those in need of adequate healthcare, can lack the knowledge to accurately report chronic conditions. Yet, as a consequence of their design, these types of studies are limited by their inability to generalize findings to survey populations. Perhaps more significantly, these inherently interpretive studies lack ability to provide measurement—a component particularly essential in the field of survey methodology where there is a requirement to account for response error in survey design.

In this paper, we demonstrate a mixed-method approach to examine the performance of two survey questions intended to measure pain. First, in-depth cognitive interviews were conducted to identify the various interpretive patterns used by respondents to formulate answers. Based on the findings of the qualitative component, a subset of structured probe-like questions was developed to quantify those interpretive patterns determine prevalence as well as to statistically compare the patterns across demographic subgroups. The questionnaire was then fielded in the 2010 National Health Interview Survey (NHIS). Analysis of NHIS data suggests that, utilized together, the two questions perform as intended. Findings also show that while respondents' sex and race did not predict false negative or false positive responses, age and education were significant predictors.

## **Who is Your Neighbor? Sampling of Migrant Populations Using Population Clustering**

Inna Becher, *Institute for Employment Research (IAB)*

In absence of high quality sampling frames for migrant populations, many special approaches have been developed. Recently, survey research literature is focused on chain-referral or network procedures. One of them - adaptive sampling (e.g. see Thompson 1996) - not only allows for cost-effective sampling but also for statistical inference from these samples.

Only little is known about the usability and properties of such adaptive procedures in the German context. Therefore, we simulate adaptive sampling for migrants in Germany based on a telephone directory containing about 31 million entries. We simulate two migrant populations of different sizes in the general population in order to evaluate properties of an adaptive sampling scheme for rare and less rare groups. Using Monte-Carlo simulations of adaptive sampling on these populations, mean and variance of survey characteristics are estimated and compared to simulations of a simple random sample (SRS) from the same sampling frame.

Results show that adaptive procedure has many advantages compared to SRS. First, adaptive samples contain about 56% more migrants than SRS of the same size. Second, estimates from adaptive samples have great precision and consistency even for rare populations. Despite the larger effectivity and high precision, there is only moderate gain in effective sample sizes for adaptive sampling compared to SRS. As such, the advantages of adaptive procedures do not lead to considerably higher costs.



## **Examining Non-Substantive Responses in China Using the Gallup World Poll**

Quan Zhou, *University of Nebraska-Lincoln*

Kay Ricci, *University of Nebraska-Lincoln*

Allan McCutcheon, *University of Nebraska-Lincoln*

Several researchers report that cultural differences influence the question-answering process (Ji et al., 2000; Haberstroh et al., 2002). Members of collectivist societies often distinguish between in-group and out-group members and have difficulty disclosing information to interviewers that are not considered to be part of their group (Smith & Bond, 1998; Johnson & Van de Vijver, 2003). Many times socially desirable responses can manifest themselves in the form of non-substantive responses, such as “don’t know” (Berinsky, 1999). Previous research suggests that groups with low power (i.e., low socioeconomic status) are thought to be more concerned with projecting a good impression, and therefore report more socially desirable behavior (Bradburn & Sudman, 1979; Harkness et al., 2003).

In this paper, we examine the characteristics of Chinese individuals who provided non-substantive responses (i.e., don’t know) to sensitive questions concerning the national government, media freedom, city leadership, local police, and financial institutions. Using data from the 2012 Gallup World Poll, a multinational probability-based survey, this study analyzes the demographic characteristics of these individuals and examines the structure of patterns of social desirability using latent class analysis. Preliminary analysis indicates these individuals are more likely to be from groups with low social power. That is, for all five items, the respondents who reported “don’t know” are more likely to be female and to have lower levels of education. Furthermore, for four of the items (excluding the item on one’s attitude toward city leadership), respondents who reported “don’t know” are more likely to have lower levels of income and live in rural areas. Given the methodological issues that often hinder cross-cultural measurement, this study focuses on providing a better understanding of non-substantive responses in collectivist cultures, and the implications for the development of survey items for cross-national research.

## **Bridging Cultural Differences Beyond Questionnaire Translation: an Exploration of Research Methodology in Cross-Cultural Research Design with Multilingual Communities**

### **Differences between Monolinguals and Bilinguals in Cognitive Processing of Survey Questions**

Hyunjoo Park, *RTI International*

Mandy Sha, *RTI International*

Gordon Willis, PhD, *National Cancer Institute, NIH*

Cognitive interviewing allows in-depth evaluation of how respondents mentally process survey questions, and has been useful in questionnaire evaluation. In recent years, the method has also been implemented in languages other than English.

The cognitive science literature had reported that monolinguals and bilinguals tend to use different areas of the brain when solving problems (Bialystok 2007; Hernandez and Bates 2001). This indicates that they are likely to encounter different problems when administered surveys, including understanding questions and selecting responses among others, even when the questionnaire is in their native language. Therefore, researchers have conducted cognitive

interviews with monolinguals when testing translated questionnaires. For example, for Chinese questionnaire pretesting, the ideal participants are monolingual Chinese speakers who speak little or no English. This way, they mirror the intended respondents to the Chinese questionnaire. However, there has been no study examining if such differences between monolinguals and bilinguals exist in a survey context despite the evident challenge to identify and recruit monolinguals partly due to their lack of social experience in the U.S.

This study will examine the number and type of problems with survey questions identified from cognitive interviews for monolinguals and bilinguals, relying on the respondent's English proficiency as a proxy measure of monolingualism. We will also investigate whether any differences in respondents' understanding of survey questions remain after controlling for demographic differences. For data analysis, we will use the cognitive interview data collected through the American Community Survey (ACS) Language Assistance Guide (LAG) project for which 258 cognitive interviews were conducted in Chinese and Korean.

This study will help survey practitioners re-evaluate the importance of recruiting monolinguals and will guide them in conducting efficient cognitive interviews.

### **School Concepts for Spanish Speaker Respondents: Improving Concept Validity in Surveys through Cognitive Interviews**

Anna Sandoval Girón, *American Institutes for Research*

Previous work on the translation of educational attainment has shown the challenges of translating concepts related to country-specific programs (Fernández, Goerman, Quiroz 2012, Schoua-Glusberg et al. 2008). In this paper I present the results of two iterative rounds of cognitive testing on a series of the National Household Education Surveys (NHES) educational questions with monolingual Spanish-speaking respondents from different countries living in the United States. The focus of the cognitive testing was on the validity of terminology such as charter schools, homeschooling, and after-school programs, with a goal of improving concept validity in Spanish. In the first round of testing the study team tested the questions using a translation that was as close as possible to the English original, in other words the translation was a literal translation of the original English. In the second round of testing the translations were revised for comprehension and re-tested. Some of these revisions included translations that were not direct and were instead equivalent to the meanings of the concepts in English, for example after school programs was initially translated as “programas extracurriculares.” On the second round it was re-translated to “actividades despues de la escuela” when respondents were not familiar with the term “extracurriculares” and felt confused by the term that appeared unfamiliar. This paper focuses on the results from the two rounds of cognitive testing; where we found that direct translations are less successful at getting reliable results and that terminology often require an explanation in order for terms to make sense in the surveys. As an increasing number of surveys seek to collect data in several languages one of the lessons learned is the importance of translations that collect reliable and valid data.

### **Adaptation of Standard Cognitive Interview Methodology for Use with Spanish-Speaking Respondents**

Patricia Goerman, *U.S. Census Bureau*

Ryan King, *U.S. Census Bureau*

There has been very little empirical research into best practices for non-English cognitive testing. However, research has found that direct translation of English language probes and

methods do not always work smoothly in non-English languages such as Chinese and Spanish (Pan, 2008; Goerman, 2006). This paper presents exploratory research on Spanish-language cognitive testing probes, introductions and instructions. Fifty cognitive interviews were conducted with Mexican-origin Spanish speakers, almost three quarters of whom had less than a high school level education with the remainder having had more than 12 years of schooling. Survey questions from the Census Bureau's American Community Survey were administered and tested simulating the Computer Assisted Personal Interview (CAPI) mode. Respondents were interviewed either using a "structured" approach or an "experimental" approach. The structured approach, which used mostly scripted probes, was designed to imitate the common situation where English-dominant researchers design a project in English and have all materials translated into the non-English language. The experimental approach involved increased rapport building and explanation of the cognitive interview process as well as rewording of traditional cognitive interview probes when they caused comprehension difficulty. Interactions between cognitive interviewers and respondents were transcribed and coded, using a modified behavior coding system. Results address whether and to what extent different types of cognitive interview probes, such as meaning oriented probes, process oriented probes, paraphrasing probes and think aloud procedures worked well with the Spanish speakers in the context of both the "structured" and the "experimental" interviews. The paper makes recommendations regarding ideal cognitive interview probe wording and study introductions and explanations to use with Spanish speakers.

### **What Does the Satisfaction with Democracy Measure Mean to Respondents in Different Countries? How Cross-National Web Probing Can Contribute to Answering this Question**

Dorothee Behr, *GESIS – Leibniz Institute for the Social Sciences*

Michael Braun, *GESIS – Leibniz Institute for the Social Sciences*

The item measuring satisfaction with democracy is a core item in (comparative) political research. However, what exactly it measures is a bone of contention among researchers (Anderson, 2002; Canache, Mondak, & Seligson, 2001; Linde & Ekman, 2003). Does it measure satisfaction with incumbent authorities, with the political system in general, or is it in fact a summary indicator of political support? Moreover, how comparable is the measurement across countries in cross-national survey research? To shed light on the measurement dimensions in a cross-national perspective, we conducted cross-national web probing: In October 2011, we fielded web surveys in Denmark, Canada, Germany, Hungary, Spain, and the U.S. About 250 respondents in each country first answered the closed satisfaction item ("And on the whole, how satisfied are you with the way democracy works in [country]?" using an 11-point-scale ranging from "extremely dissatisfied" to "extremely satisfied"). They then answered an open-ended category-selection probe asking for a reason for having chosen a particular scale value. The respondents were recruited from non-representative online panels according to socio-demographic quota. Results of the closed democracy item were compared to benchmark surveys to ensure that we were working with a reasonable response distribution. A coding scheme was developed and the answers were coded accordingly. In all countries, satisfaction with policy outcomes, incumbent authorities, and the functioning of the system in general was simultaneously measured by the item, but to different degrees across the different countries. The paper will present in detail the interpretation patterns across countries. While the results can provide a basis for further discussion among political scientists, they also allow presenting the usefulness of web probing to learn more about what we actually measure, both in national and in cross-national contexts.

## **The Impact of Question Format and Respondent Background on Data Quality in a Health Survey**

Aaron Maitland, *Westat*

A number of factors may influence the quality of the answers that respondents provide when answering survey questions. The format of the survey questions defines the task around which the respondent answers the questions. Some tasks will be relatively more difficult. A popular format, particularly in visual modes such as self-administered mail questionnaires, is the grid format where a question stem is associated with a list of items. Although these questions are often used to reduce respondent burden by grouping conceptually similar items, the resulting grids may sometimes result in questions that are more visually challenging for respondents. Furthermore, the questions in grids may be more difficult depending on the background of the respondent. Respondents with less cognitive ability or less familiarity with the survey process may have particular problems with grid questions. Preliminary evidence for these hypotheses was found in a recent data collection cycle of the Health Information National Trends Survey (HINTS). Overall, rates of item missing data were significantly higher for questions in a grid format compared to stand-alone survey questions. Questions in grids routinely had missing data rates over five percent, whereas stand-alone questions routinely had missing data rates under one percent. In addition, there were significant differences by education and language. For example, lower educated respondents routinely had missing data rates on grid questions that were two or three times higher than more highly educated respondents. This suggests that respondents with lower levels of education and lower levels of English ability are likely to have more difficulty with grid questions. This paper expands on these findings with respect to grid questions and explores some of the reasons behind why there are differences in data quality by background characteristics.

## **Practical Applications for Social Media in Survey Research**

### **Is Shooting Fish in a Barrel a Good Thing? Using Social Media and Search Engines to Recruit Respondents for Self-Administered Surveys and Pretesting**

Michael J. Stern, *NORC at the University of Chicago*

The increasing ubiquity of internet usage among all segments of US society, explosion of user generated social media sites, and internet-enabled mobile technologies have raised new questions about survey researchers' ability to sample from web. As a result, there is a need to investigate whether researchers can effectively sample from the web and, if so, what is the quality of the data produced? In this talk, I address these questions by comparing and evaluating two different forms of opt-in web sampling experiments. The first sampling methods involve using advertisements as survey invitations in two different places on the web: Facebook and Google. In both sampling approaches, respondents click on the invitation advertisement posted in the banners and side-panels and are taken to our landing page with information about the 21-item survey and its entry point. Here I assess the 1) data representativeness using the General Social Survey as our national benchmark, 2) time taken to reach completes by strategy, and 3) the cost efficiency of these sampling strategies. In a second set of experiments, I explore the use of single-item Google Consumer surveys for comparison of results from alternate forms of two questions (with two different versions for each) that were tested with cognitive interview techniques. Historically, researchers have pretested questions with cognitive interviewing in various incarnations (e.g., concurrent think-aloud, retrospective think-aloud, focus group discussions, probes, and memory cues) and by embedding experimental treatments in web

surveys among undergraduate students have predominated. As a result, if a researcher wanted to assess a single-item among a large heterogeneous audience, their options were limited. In all, these experiments begin to shed light on the effectiveness of recruiting from the internet. I will summarize what we have learned, challenges experienced, strengths and limitations, and future research that should be conducted.

## **A Virtual Focus Group? Using Social Media Sentiment Analysis Techniques to Improve Questionnaire Design**

Christine Pierce, *The Nielsen Company*

Social Media analysis techniques are commonly used in market research and increasingly more in public opinion research. This paper seeks to apply social media sentiment analysis techniques to improve the design of the questionnaire and materials for an ABS-frame TV Diary survey. In performing this analysis, we searched popular social media websites such as Facebook and Twitter as well as blogs and posts from other user-generated content sites. We found an extensive amount of feedback from respondents who had received the TV Diary and recruitment materials. We analyzed the respondent feedback (over 800 posts) using text and other sentiment analysis techniques. The analysis will display the respondent feedback segmented into positive, negative, and neutral. It will discuss how we used the feedback to help understand the respondent perception of the diary keeping experience. While admittedly a self-selected group of respondents, the results helped identify areas of the survey and recruitment process which were burdensome or confusing. Ultimately, the results of the social media analysis will be combined with results from more traditional content testing techniques to redesign the content and design of our survey materials. We explored different analytic methods and discuss the benefits and drawbacks for different approaches in performing this type of research.

## **Contacting Sample Members by Facebook or Email: What Works?**

Bryan Rhodes, *RTI International*

Social networking sites like Facebook have been shown to provide effective tools for locating and contacting sample members in longitudinal surveys. However, locating sample members on social networking sites creates unique challenges. Sample members with common names or private profiles can be difficult to locate. Since email addresses are typically unique to an individual and are often linked to a social networking profile, having a sample member's email address may be particularly helpful in locating. However, if a study has access to a sample member's email address, what advantage does contacting the sample member via a social network have over just sending a traditional email (which is often less labor intensive)? Some studies have found that the use of email for everyday communication is dropping, and is often being replaced by communication via social networking sites. This is particularly true among younger people. This study compares contact via email and Facebook to determine which method is most effective in which situations.

We conducted an experiment as part of the SEED for Oklahoma Kids (SEED OK) survey, a longitudinal survey of new mothers about saving for college. As part of the SEED OK study we conduct sample maintenance between survey waves. During sample maintenance, we attempt to contact all sample members and ask them to update or confirm their contact information (address, phone, etc.). For this experiment we randomly assigned the sample to either be contacted by email or through Facebook. We examine the number of sample members who update their contact information from each group. Since email or Facebook contacting may be

more or less effective among different groups, we also explore the demographic and other differences between the responders.

### **Reporting Data in the Social Media World**

Casey Tesfaye, *American Institute of Physics*

In the fall of 2013 the persistently low proportion of women in physics became a hot topic of discussion. This discussion was in part fueled by a series of reports that grew additional lives once they were published, echoing among individual blog posts, news sites such as the Huffington Post and the New York Times, and reverberating into scientific magazines. The published pieces generated countless comments and tweets among readers and interested parties. What follows is an analysis of the interpretations, misinterpretations and trajectories of published findings, anchored in the “women physics” keyword search on Twitter. This goal of this analysis is to generate helpful, data-driven guidelines for researchers who are now forced into a social media space and hoping to stem misinformation and misreports of their results.

### **The Sociality Hierarchy and the Future of Social Media and Survey Research**

Craig A. Hill, *RTI International*

Survey researchers have dedicated their professional lives to producing high-quality, fit-for-use data—data that are accurate, timely, and accessible. Can we reach these same outcomes or goals using social media data? The coming Age of Ubiquity will provide social science researchers with “Big” (organic) Data, produced by multiple sources (including social media). Can we arrive at a deeper understanding of these data, so that we understand their sources, their limitations, and their potential?

In this discussion, we’ll describe a categorization scheme for social media data (the ‘sociality hierarchy’) and talk about ways that social media data are already being used to supplement survey data, and about possible ways in which social media data can be used in the near-term future, and in the longer term future. The sociality hierarchy has three levels: broadcast, conversational, and community. At each level of the hierarchy, social science researchers should consider how the nature of the communication platform influences their choices about how to employ social media platforms.

## **Strategies for RARE Populations**

### **Locating Rare Populations Using Social Media: Twitter and Persons Displaced by Hurricane Sandy**

Martin Barron, *NORC at the University of Chicago*

Trevor Thompson, *NORC at the University of Chicago*

Jennifer Benz, *NORC at the University of Chicago*

Locating rare populations for surveys pose a difficult and costly challenge for survey researchers. One relatively new source of potential information on certain rare populations can be gathered using social media. In this paper we investigate the potential for using social media to locate one rare population.

This study examines the feasibility of using Twitter to locate persons displaced by Hurricane Sandy. Using the geographic information embedded in (some) Twitter posts, we identified users

who were posting in Hurricane Sandy affected areas prior to the hurricane and from other locations after the hurricane. These potentially displaced individuals were then invited to participate in an online survey of hurricane experiences. Among other things, the survey is used to confirm our assumption that the identified individuals were displaced.

Our paper reports on findings of our locating and recruitment effort. In addition to reporting on the number of identified displaced persons and our success at recruiting Twitter users to participate in an online survey, we also discuss the limits of our data for making inferences to the larger population. We conclude with some thoughts on how social media might be used with rare population sampling methods for making inferences to rare populations.

## **Effective Survey Sampling of Rare Subgroups: Probability-Based Sampling Using Split-Frames with Listed Households**

Mansour Fahimi, *GfK*

Targeting rare subgroups via traditional survey sampling methods, such as address-based sampling and random digit dialing, can be cost-prohibitive because the standard application of such methods can require extensive field resources for screening purposes. In contrast, complete reliance on targeted lists that can secure higher hit-rates can seriously undermine the integrity of the resulting samples because meaningful selection probabilities cannot be calculated from incomplete sampling frames. However, augmenting an incomplete frame with the remaining sampling units in the domain of interest can create a complete frame from which probability-based samples could be selected. Specifically, by relying on disproportionate stratified sampling techniques representative samples can be produced that will significantly reduce screening costs. The author will present a comprehensive strategy for split-frame sampling methodology, including a simple optimization technique for determining the optimal allocation of the total sample through minimization of the overall design effect.

## **Efficient Dynamic Venue-Based Sampling Designs for Hard-to-Reach Populations**

Ronaldo Iachan, *ICF International*

Tonja Kyle, *ICF International*

This research presents state-of-the-art venue-based and real-time sampling methodologies drawn from a few recent studies designed, conducted and/or managed by ICF. The studies seek to survey hard-to-reach populations of men who have sex with men (MSM) and of minority urban teens. The approaches have been refined to balance statistical efficiency with practical considerations. This paper examines empirical findings from the studies that inform design refinements.

We begin with an overview of the sampling designs developed for the studies. We describe a dynamic venue-day-time sampling (VDTS) system developed by ICF. The system allows for building monthly sampling frames of venues and venue-day-time (VDT) as the basis for a multistage sampling of visits and individuals. Then, we provide an overview of the complex weighting procedures that account for the differential sampling probabilities and nonresponse propensities, and to adjust for the multiple opportunities of selection for individuals who attend the venues on the frame. The discussion addresses the challenges in computing proper probabilities and adjustments based on the survey data. We examine inefficiencies that may arise from weight variability, and quantify their impacts on design effects and variances for key estimates.

We provide recommendations designed to mitigate these limitations through improvements in the study design and procedures. Upcoming study cycles will collect data that will support analyses comparing alternative designs. Design changes include standardization for constructing venue and VDT frames, and collecting measures of size that will lead to a more efficient sampling design.

## **Utilizing Cell Phone Rate Center Information in Targeting Low Income Populations**

Eran N. Ben-Porath, *SSRS/Social Science Research Solutions*

David Dutwin, *SSRS/Social Science Research Solutions*

Mollyann Brodie, *The Kaiser Family Foundation*

How efficient are cell phone rate centers as a means of targeting uninsured and low-income populations? The rapid decline in coverage of landline sampling frames, particularly among younger Americans, poses a unique problem in targeting these harder-to-reach populations. Whereas Census-based demographics can be associated with landline telephone exchanges, thus allowing survey researchers to target certain populations such as those in lower-income households, similar methods for stratifying cell phone frames are relatively new and untested. In this paper, we present data from two large-scale studies with sizable cell phone components aimed at reaching hard-to-reach populations such as the uninsured, and those with very-low family incomes (under 138% of the federal poverty level). The results from these two studies are encouraging. Cell phone stratification, based on billing centers rather than exchanges, performed as efficiently as the landline design. Incidence of uninsurance in California rate centers associated with lower-income areas was almost twice as high as in assumed higher-income areas. Similarly, the results from a multi-state study targeting low-income populations, found a strong monotonous relationship between the estimated share of low-income populations across five sampling strata and the actual income distribution observed in the field.

This paper discusses the methodology, analyzes the findings based on more than 25,000 screens, and discusses the implications and limitations for future studies of hard-to-reach populations.

## **Targeting Hard-to-Reach, Lower SES Respondents in an Experimental ABS Design**

Robyn Rapoport, *SSRS/Social Science Research Solutions*

Rebekah Gould, *Oregon Health Authority*

David Dutwin, *SSRS/Social Science Research Solutions*

Address-based sample (ABS) designs allow for the best possible coverage of the general population since virtually all of the target population elements are included in the Delivery Sequence File sample frame. At the same time, research points to the potential for disproportionate underrepresentation of lower SES households for which a listed telephone number is not available (unmatched sample) in ABS designs offering multiple modes to complete the survey.

As part of the 2013 state-wide Oregon Health Insurance Survey, SSRS partnered with the Oregon Health Authority in a two-pronged approach aimed at determining how ABS designs can be enhanced to increase representativeness among the final sample of completed interviews.

First, we tested the following experimental conditions:



1. Sending a \$2 pre-incentive to 20% of matched and unmatched households.
2. Offering either a \$10 or a \$20 post incentive to unmatched households.
3. Designating 10% of the matched and 15% of unmatched sample to complete the survey via telephone or web but not hardcopy; while all three modes were available to others in the sample.

Second, we included questions asking respondents if they were aware of the different modes they could have used to complete the survey and why they chose to complete the survey online, by telephone, or via hard copy.

Preliminary data analyses suggest that including a pre-incentive is correlated with a disproportionately larger increase in completed interviews from unmatched sample than matched sample. In addition, answers to the respondent questions indicate that the vast majority of those who completed the web and hard copy instruments or called in to complete the survey were aware of other options while just over one third of respondents reached via an outgoing call were aware of other options for completing the survey.

## **Questionnaire Design: Visual Design Effects and Data Quality**

### **Impact of Response Scale Direction on Survey Responses in Web and Mobile Web Surveys**

Florian Keusch, *University of Michigan*

Ting Yan, *University of Michigan*

Saram Han, *University of Michigan*

Lirui He, *University of Michigan*

Due to the rising penetration rate of smart phones, web surveys are increasingly accessed and completed on a smart phone or tablet instead of a personal computer. In addition, mobile web surveys are gaining popularity as an alternative self-administered data collection mode. However, the limited amount of space for displaying questions and scales on mobile screens poses new challenges for survey researchers. One of them is the optimal design of response scales. Survey literature has shown that design features of response scales affect survey responses and that answers to rating scales differ between web and mobile web respondents. This paper focuses on one design feature of a response scale – the direction of scale – and studies the impact of the direction of response scale on the actual answers. This paper reports a methodological experiment crossing three factors (mode of administration, scale direction, scale alignment) in a full factorial design. Half of the sampled are invited to complete a web survey on their personal computer and the other half are invited to fill out the same questionnaire on their smartphone. All rating scales are shown to run either from positive to negative (or high to low) or from negative to positive (or low to high). In addition, all response scales are presented either horizontally, vertically, or as a drop-down menu. Drawing on empirical findings from web studies, we hypothesize a stronger scale direction effect among mobile web respondents than PC web respondents.

## **The Whole is More than the Sum of its Parts: Understanding Item Nonresponse in Self-Administered Surveys**

Alian Kasabian, *University of Nebraska-Lincoln*

Jolene Smyth, *University of Nebraska-Lincoln*

Kristen Olson, *University of Nebraska-Lincoln*

Item nonresponse is typically either treated as the result of respondents who lack motivation or cognitive ability (Krosnick 1991) or a breakdown of the cognitive response process driven by characteristics of the questions themselves (Beatty and Hermann 2002). In this paper we examine predictors of item nonresponse in a self-administered survey. Like previous research, we include respondent (age, education, etc.) and question (reading level, question format, question type, etc.) characteristics as predictors, but unlike much previous research, we also examine the effects of question and questionnaire visual design characteristics (location on the page, headings, page format etc.) on item nonresponse. While some visual design literature has examined how features of individual questions impact item nonresponse, little research has examined the effects of bigger layout decisions (i.e., arranging questions vis-à-vis one another) that are made in the design of every questionnaire. Despite the lack of empirical research on this topic, visual design theory suggests that different features of the questionnaire layout can make items more likely to be noticed, which is a prerequisite for response (Jenkins & Dillman 1997; Dillman et al. 2009). In addition, from the current literature, it is unclear how respondent, question, and questionnaire layout characteristics jointly affect item nonresponse rates. We utilize cross-classified random effects models to predict item nonresponse using two mail surveys of Nebraska adults (AAPOR RR1=27.2%, n=959, 207 questions; AAPOR RR2, n=565, 98 questions). Preliminary results indicate that question and questionnaire characteristics explain 3-5 times more variance in item nonresponse than respondent characteristics. Moreover, several of the questionnaire features typically associated with increased item nonresponse (e.g. sensitive questions, grids) when examined in isolation failed to gain significance when modeled alongside respondent and questionnaire characteristics. These findings suggest a more holistic understanding of the effects of questionnaire design on item nonresponse is needed.

## **The Effect of CATI Questionnaire Design Features on Response Timing**

Kristen Olson, *University of Nebraska-Lincoln*

Jolene Smyth, *University of Nebraska-Lincoln*

Recent evaluations of questionnaire design features have focused on self-administered questionnaires (Dillman, Smyth and Christian 2009). Less consideration has been paid to how the layout and format of interviewer-administered questionnaires affects the time to ask and answer survey questions. This paper looks at the design of screens in two CATI surveys, the Work and Leisure Today survey (n=450, AAPOR RR1=4.2%) and a subset of Gallup's Private Wealth Survey (n=450). We examine the effect of question and questionnaire characteristics on response time as measured by paradata using cross-classified random effects models (Couper and Kreuter 2013; Yan and Tourangeau 2008). We join these with systematic coding of the characteristics of each screen and question. We draw on previous research that indicates that fast question answering indicates better quality for some question characteristics, but lower quality for others (Olson and Parkhurst 2013). Preliminary findings indicate that questions that include definitions take significantly longer to administer and answer ( $p=0.07$ ), questions in a battery are significantly faster ( $p<.001$ ), and that questions asked over two screens take significantly longer ( $p<.001$ ). We examine characteristics of questions that have received empirical attention, including the format and type of question, the number of response options,

and labeling of scale points. We also examine other features that are common to questionnaire design, but have received less empirical attention in telephone surveys, including the use of special characters (such as parentheses, quotation marks, and dashes) and a question's reading level. This analytic approach allows us to examine multiple questionnaire design features simultaneously and their effects on response latencies. Implications for questionnaire design are discussed, including how individual question and questionnaire features are affected by the context in which they appear and how this affects the time to ask and answer survey questions.

## **Improving Federal Forms with User Experience Testing and Eye Tracking**

Jennifer Romano Bergstrom, *Fors Marsh Group*

Jonathan Strohl, *Fors Marsh Group*

Andrew Hale, *Fors Marsh Group*

Sarah Keaton, *Fors Marsh Group*

Many modern government forms have poor usability. Usability testing ensures that forms use plain language that is comprehensible to consumers, contain clear format and design, effectively communicate information to consumers, and are feasible across various mediums used by consumers. In usability testing, representative participants complete tasks while the researcher notes difficulties participants encounter. Eye tracking is a unique technology that can be used during usability testing that captures participants' eye movements while they work through a survey or form. Above and beyond the performance data obtained during usability testing, eye tracking can inform about how consumers process forms and surveys.

We present findings from eye-tracking-based usability studies of federal forms (paper and electronic) that included long pages of instructions, as well as supplementary documents required to properly complete the forms. We describe several methods that we have used to evaluate government forms, and we demonstrate common issues that we frequently discover, such as (1) organization of instructions do not promote reading; (2) jargon language does not conform to everyday language and is confusing; (3) a lack of usable information in fillable sections leads to issues with completing the forms. Eye tracking provides a deeper level of analysis than traditional usability metrics alone. Implications and ways to prevent these issues will be discussed.

## **Examining the Use of Visual Grouping to Reduce Skip Pattern Errors**

Rebecca J. Powell, *University of Nebraska-Lincoln*

Many survey researchers are implementing mail surveys since they can achieve better coverage (Iannacchione, 2011) and response at a lower cost than with telephone surveys. However, tradeoffs occur since respondents need to correctly navigate skip patterns without computer or interviewer assistance in order to avoid incorrect or missing data. Previous research by Redline and colleagues (2003) aimed to determine the best visual design of skip instructions to decrease skip errors. They implemented bolded instructions, arrows to direct respondents, and verbal feedback statements in follow-up questions (e.g., if yes,) in their best method, however, 17.5% of respondents still made errors. This research expands on their work by testing whether implementing grouping principles from the Gestalt laws of pattern perception (Ware, 2004) to delineate between items that everyone should answer and items that only certain people should answer can further reduce skip errors. Using a 2x2x2 factorial experimental design, this study examines the effectiveness of using separate enclosures, indentation, and outline style numbering on the reduction of errors. Specifically, fewer skip

errors are expected when follow-up items are enclosed in a separate region from the main questions versus in the same region as the main questions (principle of enclosure), indented relative to the main questions versus aligned at the same margin (principle of proximity), or distinguished by sub-numbering (e.g., 2a, 2b, etc.) rather than numbered in the same style as the main questions (principle of similarity). Furthermore, this study examines the combined effect of these visual grouping strategies on skip errors. These grouping principles are tested within a mail survey about labor force participation sent to 12,000 Northeast Nebraska residents. The survey was fielded from October through December 2013.

## **Satisficing, Response Burden and Data Quality**

### **Instant Interactive Feedback in Grid Questions: Reminding Web Survey**

Respondents of Speeding and Nondifferentiation

Tanja Kunz, *Darmstadt University of Technology*

Marek Fuchs, *Darmstadt University of Technology*

Interactive Web-based surveys permit direct interventions depending on respondents' behaviors similar to interviewer-administered surveys. Previous findings on interactive feedback addressing speeding and nondifferentiation in grid questions showed that prompts slowed down response times and reduced straight-lining in later grid questions without increasing the risk of survey breakoff (Conrad et al., 2009, 2011; Zhang, 2012, 2013). Although these findings are promising, effects of interactive feedback on the prevention of satisficing behaviors in grid questions are rather small. This may be due to the fact that these studies used either a pop-up window or a follow-up Web page to deliver the interactive feedback to respondents. Thus, respondents were supposed to actively return to the grid affected by speeding or nondifferentiation. By contrast, the present experiments implemented interactive feedback that was instantly triggered depending on respondents' satisficing behaviors, i.e., speeding and nondifferentiation while the respondents were still answering a particular grid question. The underlying idea was that feedback provided instantly while the items were still being answered can address satisficing behaviors more efficiently as compared to delayed feedback provided after the grid was submitted.

In a randomized field experimental Web survey conducted among study applicants (n=7454), a between-subjects design was implemented to test the effect of instant interactive feedback with two different kinds of messages asking respondents to either take more time to answer (experiment 1) or to differentiate more between responses to items of the same grid (experiment 2). Initial findings indicated that speeding and nondifferentiation were significantly reduced by either feedback message. Furthermore, effects carried over to later questions. These results suggest that instant interactive feedback has the potential to induce immediate and sustainable improvements in respondent behavior. In addition, results are discussed in view of the precise timing of feedback to obtain best possible effects concerning improved data quality.

### **Respondents Playing Fast and Loose?: Antecedents and Consequences of Respondent Speed of Completion**

Randall K. Thomas, *GfK Custom Research, LLC*

Frances M. Barlas, *GfK Custom Research, LLC*

Paradata can be used to examine respondent behaviors such as speed of survey completion, extent of response differentiation, failure to follow directions, and response inconsistency.

Paradata can enable researchers to either alert and re-engage respondents in an effort to counter their reduced attention and increase their response quality or exclude them so they do not count as qualified completes. Using the Advertising Research Foundation's Foundations of Quality 2 dataset that had over 57,000 respondents from 17 different providers using opt-in panel members, we examined the impact of speed of completion on data quality and other associated sub-optimal respondent behaviors (e.g., response non-differentiation and response inconsistency). Although, we found that speeding was associated with other forms of sub-optimal behavior along with demographic factors in predictable ways, those who sped or exhibited other forms of sub-optimal behavior had very little impact on overall results because the number of respondents exhibiting the worst of the behaviors tended to be small.

## **Response Burden: What Predicts It and Who are Burdened Out?**

Scott Fricker, *U.S. Bureau of Labor Statistics*

Ting Yan, *Institute for Social Research, University of Michigan*

Shirley Tsai, *U.S. Bureau of Labor Statistics*

Concerns about the burden that surveys place on respondents have a long history in the survey field. As early as the 1920s, survey researchers and organizations were warned about response burden. However, a review of the survey literature reveals gaps in research on response burden – undeveloped conceptualization of response burden, the lack of good measurement of the subjective perception of response burden, and the lack of empirical research on what predicts response burden and what is the impact of response burden on data quality. The Consumer Expenditure Interview Survey (CE) includes survey questions to measure respondents' perception of response burden and their attitudes and perceptions about the survey. Using CE data, this paper attempts to advance the research on response burden by examining both survey characteristics and respondent attitudes and beliefs that affect respondents' perception of burden. The objective of this research is to identify respondents who are at a greater danger of feeling burdened and the circumstances under which they are more likely to be burdened out. We will also quantify the impact of response burden on data quality. The results of this research will advance our understanding of the causes of response burden, and identify the missing elements in the current measurements of response burden, and under the potential impact of response burden on data quality.

## **The Effects of Splitting Long Surveys Into Two**

Annie Pettit, *Peanut Labs*

Despite the fact that response rates decrease every year, researchers continue to demand that responders answer long surveys. We know that response rates are higher for short surveys but we continue to demand more effort from responders. We feel that we cannot afford to run two separate surveys and we cannot action data that comes from two different groups of responders.

In this presentation, we'll demonstrate the differences between results from one long survey vs two short surveys. Test group 1 will respond to a standard survey with a standard incentive. Test group 2 will respond to the first half of the survey, and test group 3 will respond to the second half of the survey. The two spit-surveys will include sufficient duplicate questions to permit appropriate analyses, and the incentive will match the length of the half-survey relative to the full-survey.

Three main features will be examined in this study.

1. First, we'll examine differences in data quality such as random responding, straightlining, underclicking, speeding, red herrings, and more. Will the short survey exhibit fewer poor responding behaviours?
2. After removing poor quality data, we'll also look at the magnitude of differences in terms of raw results, rankings, mean scores, and research conclusions. Will the two short surveys provide essentially the same conclusions as the long survey?
3. Finally, we'll examine the responder's enjoyment ratings of the surveys. If the two short surveys generate more positive responder scores as well as similar or better data quality, and similar conclusions, then researchers will have no alternative but to seriously attempt to run more split-surveys.

**Friday, May 16**  
**10:00 a.m. – 11:30 p.m.**  
**Concurrent Session D**

## **Data Collection Through Mobile and Electronic Devices**

### **Mixed-Devices in a Probability Based Panel Survey. Effects on Survey Measurement Error**

Peter Lugtig, *Utrecht University, Department of Methods and Statistics*

Vera Toepoel, *Utrecht University, Department of Methods and Statistics*

Recent studies have indicated that members of Internet panel surveys now complete surveys on PCs, but also on subnotebooks, tablets or smartphones (Bosnjak et al 2013). This means that web surveys are displayed on very small and very large screens, and are completed with keyboard, mouse or touchscreen.

The fact that Internet surveys can now be completed with so many different devices may have implications for measurement. There have been only a few studies studying the effects on measurement error between different devices. Most studies have concentrated on comparing mobile phones to desktop PCs and have found no clear differences between them with regards to measurement error. Because respondents always self-select into particular devices, it is very difficult to separate selection effects from measurement effects of the device however.

This paper aims to fill this gap by using longitudinal data from a large Dutch probability-based panel survey (LISS) to analyse over time 1) whether the same respondents use different devices to complete Internet surveys and 2) assess whether a device switch over time affects survey measurement error. To this end, we coded the User Agent Strings of 6 waves of data, and categorized respondents according to the devices used over time. We identified several groups of 'stable'-device users (e.g. "always PC") and 'switchers' (e.g. PC -> tablet -> tablet -> PC).

We study the measurement properties of each of the devices using a within-subjects design. We look at 1) the number of Don't knows and item missing 2) interview duration 3) straightlining 4) the length of open answers 5) primacy effect in check-all-that-apply questions and 6) how the survey is evaluated at each wave.

We conclude with a discussion of the effects our findings have for the design of web surveys.

### **Recruitment and Retention of a Passive Mobile Panel**

Robert DeHaan, *The Nielsen Company*

Kim Main, *The Nielsen Company*

What is the most effective approach to recruit panelists into a panel that passively measures their mobile activity via a downloaded app? Once you recruit respondents to the panel, will they keep the app on their phone?

In Spring 2013, a feasibility study was conducted to evaluate the effectiveness of recruitment via different modes (telephone placement calling, email solicitation to a web recruitment survey, and mailed letters with a link to a web recruitment survey). Once potential panelists had consented to participate during recruitment, they were sent a link to download a mobile application that passively collects their mobile usage data. Panelists were not incentivized during the recruitment phase, but were paid for each month their mobile application reported activity. The research questions for this study include:

- What would cause telephone placement calling to perform much more poorly than emails and mailed letters?
- How did the attributes of participants that completed the download compare to those who did not participate (no contact, declined, agreed but did not download)?
- What were attrition rates during the 3 month data collection phase and were these panelists (or their phones/usage) systematically different than retained panelists?
- At the end of the 3 month study, after cancellation letters were sent notifying participants payments would cease, would people continue to passively participate in the panel?

In this presentation, we will present the results from the feasibility study. We will review the overall recruitment/download rates as well as present a cost model analysis of the various recruitment methods.

## **Use of Smartphones to Collect Information about Health Behaviors: A Feasibility Study**

Sean Hu, *Centers for Disease Control and Prevention*

Naomi Freedner, *ICF International*

Piper DuBray, *ICF International*

Shanta R. Dube, *Georgia State University, School of Public Health*

The evolution of mobile communications technologies, such as smartphones, provides a unique opportunity for innovation in public health surveillance. Web access via smartphone is immediate, accessible, and confidential - a combination of features that could make it a prime candidate for the ongoing research, surveillance, and evaluation of risk behaviors and health conditions. The purpose of this pilot study was to conduct process and outcome evaluations of a web-based follow-up survey that used smartphones to collect population-based data. The study consisted of a baseline telephone interview with respondents aged 18 to 65 drawn from a national sample frame of random-digit-dial cell phone numbers. Respondents were asked a short series of demographic questions, general health questions, and questions about tobacco and alcohol use. One and two weeks after completing the baseline interview, respondents received a unique link via text message to complete a web-based follow-up survey. The study was designed to explore whether and how smartphones can be used to collect population-based public health and behavior data. Findings from this pilot study will be presented to: 1) provide insights into the feasibility of using mobile devices in public health surveillance; 2) address methodological issues such as response bias, coverage bias, data quality, and costs; and 3) present preliminary outcomes. As mobile communications continue to evolve, a better understanding of how smartphones can be used to collect data on risk behaviors and health conditions is critical to public health surveillance and evaluation efforts, including rapid response, panel, or longitudinal studies.



## **Preferences and Attitudes Toward Use of Electronic Devices for the 2020 Census**

Mandy Sha, *RTI International*

Emily Geisen, *RTI International*

Murrey Olmsted, *RTI International*

Elizabeth Nichols, *U.S. Census Bureau*

Tim Flanigan, *RTI International*

Georgina McAvinchey, *RTI International*

Sabin Lakhe, *U.S. Census Bureau*

Survey researchers often strive to incorporate emerging technologies into their methodology to entice and facilitate respondents' participation. However, respondents are rarely asked beforehand about their preferences for the use of technology, either by interviewers or for completing self-administered surveys. This paper provides a qualitative picture and a snapshot in time about respondents' perceptions, preferences, and acceptance levels in using mobile devices (smartphones and tablets), laptops/desktops, and paper forms to respond to government surveys.

As the largest survey in the United States, the Decennial Census has been fielded using paper questionnaires. Research is underway to develop the 2020 Census questionnaire for use on electronic devices that can be either self- or interviewer-administered. In this cognitive interview study, participants were assigned to complete the Census questionnaire using an electronic device or paper, followed by a debriefing interview. Participants were asked about their mode preference for completing the Census questionnaire by themselves or with an enumerator, and whether they owned an electronic device. While most participants were comfortable with using electronic devices to complete the Census, they expressed concerns about using a smartphone—even when they owned or regularly used this device. However, support for smartphones was higher when the device was to be used by an interviewer, not the participant. In contrast, the perception of tablets was quite favorable, even though ownership of tablets was much lower than smartphones among the participants. When examining the participants' demographics, age was the most significant factor related to device preference. As electronic technologies become more accessible to the public, this information may inform survey practitioners about respondent motivation for and barriers to participation in the use of electronic devices to respond to government surveys.

## **The Generational Technology Divide and Implications for Smartphone Data Collection**

Michael W. Link, *The Nielsen Company*

Shu Duan, *The Nielsen Company*

Kelly Bristol, *The Nielsen Company*

Jennie Lai, *Independent Consultant*

Smartphones and their expanding capabilities have the potential to revolutionize the collection of attitudinal and behavioral data. Survey capability, GPS location tracking, picture and video capture, and study-specific applications can facilitate the collection of a rich array of information and facilitate in-the-moment data capture. Contrary to popular wisdom, however, not everyone has a smartphone which can result in coverage errors unless phones are provided, but perhaps more importantly not everyone knows how to use or feels comfortable using the many features of a smartphone. This is particularly true of "older" adults (including some in their 40s and 50s) and/or those with less experience using these technologies. We report the results of a study of persons 15 and older which utilized a two-week activity diary to capture television viewing

behaviors where respondents could choose to use a smartphone, tablet, or website -- or any combination of these during the field period. The data show that respondents 50 years of age and older were far more likely to choose to respond via a website than a mobile device. In doing so, however, they were also more likely to make their entries retrospectively (not the same day as the viewing occurred) hence increasing the chance of recall error and to record fewer hours of viewing than younger adults (the reverse of what other methodologies have shown) thereby indicating a degree of measurement error. In both in-person and survey study debriefs, a higher percentage of older adults also indicated they had difficulties using the smartphone device. These results have significant implications for the broader use of smartphones as data collection tools both for the general public and in particular for older adults.

## **Quienes Somos: Racial Identification, Social Conservatism and Religious Affiliation**

### **Evaluating Race and Hispanic Origin in Administrative Records Relative to the 2010 Census and 2010 American Community Survey**

Sonya Rastogi, *U.S. Census Bureau*

James Noon, *U.S. Census Bureau*

Renuka Bhaskar, *U.S. Census Bureau*

Ellen Zapata, *U.S. Census Bureau*

Administrative records could be used to reduce costs, reduce respondent burden, and improve data quality. One potential use of administrative records is to use the data when race and Hispanic origin responses are missing. Previous research, largely evaluating administrative records relative to Census 2000, has indicated that race and Hispanic origin responses for minorities in administrative records are of lower quality relative to non-Hispanic Whites. This paper, building on research from 2000, seeks to better understand race and Hispanic origin data in administrative records relative to the 2010 Census and 2010 American Community Survey (ACS). This paper discusses methodologies for assigning race and Hispanic origin to one record when these responses are divergent across administrative records. Then, using multinomial regression, it assesses the characteristics associated with responses that match between administrative records and the 2010 Census and 2010 ACS, do not match, or are missing in administrative records. Preliminary findings from research between administrative records and the 2010 Census suggests that administrative records race and Hispanic origin responses have higher agreement for non-Hispanic Whites relative to other race and ethnic groups. Some of the findings for specific race and ethnic groups can be situated in survey methodology and sociological literatures. For example for Hispanics, questionnaire design and conceptual complexities between race and Hispanic origin likely contribute to non-matching and missing responses in administrative records. Also, racial identification and reporting can be fluid - either situational or change over one's life course. Our results indicate that racial fluidity may be occurring for populations who have large proportions of people who identify with their mixed heritage – such as American Indian and Alaska Natives and Native Hawaiians and Other Pacific Islanders. This research will help us understand data quality issues to inform future data partnerships, acquisitions, and also target fieldwork.

## **Half the Story: Hispanic Representation on Online Panels**

Jordon Peugh, *SSRS/Social Science Research Solutions*

David Dutwin, *SSRS/Social Science Research Solutions*

Hispanics in the U.S. have increased by 43% over the last decade and there is increasing interest in understanding their opinions and experiences. However, this is a costly population to reach for survey research for three reasons: prevalence (while growing, Hispanics still represent only 14% of all adults in the U.S.); language (29% of Hispanics do not speak English well); and accessibility (over half of Hispanics are not reachable on a landline telephone and 35% do not have Internet access at home).

Many Hispanic population researchers have considered or utilized online panels as a low cost way to reach this group. The question we explore in this paper is the extent to which typical online panels represent the full Hispanic population.

Looking at data collected by telephone using dual frame RDD sample and bilingual interviewing, we find that just over half of Hispanics (55%) could possibly be represented on a typical English-language convenience web panel, meaning they can read in English at least “pretty well” and use the Internet at least occasionally. This leaves 45% of the Hispanic population completely missing from most online panel samples.

We looked further at these Hispanic respondents who could be included in an online panel and found dramatic differences from their counterparts. Namely, potential web panel members are highly acculturated. They are more than twice as likely to be U.S. citizens; six times as likely to have a college degree; nearly twice as likely to have a driver's license; more than twice as likely to use social media; and more than four times as likely to consume TV news in English, among other stark differences. This calls into question the usefulness of typical web panels for researchers of the Hispanic population. Our paper will further detail differences between these two groups and implications.

## **How Socially Conservative Are Latinos? The Shifting Landscape of Latino Opinion**

Carolyn Funk, *Pew Research Center*

Jessica H. Martinez, *Pew Research Center*

Mark H. Lopez, *Pew Research Center*

Some argue that conservatism among U.S. Latinos on social and family issues creates a window for Republicans to align the Latino vote with the GOP. We draw on a rich series of representative surveys of Hispanics conducted by the Pew Research Center, 2002 through 2013, to examine the extent to which Latinos hold socially conservative views and the degree to which those views have changed over time. The findings show that Latinos tend to be more conservative than the U.S. general public when it comes to abortion. But, there are either no, or surprisingly modest, differences between Latinos and the general public when it comes to attitudes on other social issues, such as same-sex marriage, or family and gender roles. Further, there are strong opinion divides among Latinos which reflect several broad changes underway in the U.S. Latino population including the growing share of U.S. born second and third generation immigrants, the growing share of Latinos from other Central and South American countries in addition to Mexico, and the shifting religious profile of Latinos from predominantly Catholic to one with more religious pluralism including a growing share of religiously unaffiliated. We use multivariate analyses to test for the independent effect of nativity,

heritage and religious affiliation after controlling for factors such as gender, age, language-dominance and socioeconomic background. Together, these findings underscore the changing landscape of Latino opinion when it comes to social issues; we discuss the political implications of these findings.

### **The Generation Gap: Social and Religious Change Among Latinos**

Carolyn Funk, *Pew Research Center*

Jessica H. Martinez, *Pew Research Center*

The size of the Latino population in the U.S. has grown roughly four-fold since 1960 and is projected to roughly double from 16% in 2010 to 31%, by the year 2060, according to the U.S. Census Bureau. There is, also, increasing diversity among Latinos in the U.S. This paper will focus on the changing religious profile among U.S. Latinos with special attention to differences among Latinos by immigrant generation (first-generation, foreign born Latinos compared with U.S. born second and third generation immigrants). The analysis will focus on a new representative survey of about 5,000 Latinos conducted by the Pew Research Center in 2013 and will draw from other Pew Research Center surveys of Latinos conducted since 2006. The findings show dramatic changes taking place in religious affiliation with the share of Catholics declining and an increasing number of religiously unaffiliated, especially among younger, U.S. born Latinos. The result is a hollowing out of the religious middle anchored by highly observant, renewalist Christians on one end and the religiously unaffiliated on the other end. Further, there is a sizeable generation gap in behaviors and preferences for an ethnic church, offering worship services in Spanish with predominantly Hispanic congregants. The multivariate analyses also test for the independent effect of nativity in these patterns after controlling for factors such as gender, age, heritage (e.g., Mexican, Puerto Rican, Cuban, Dominican and Salvadoran), language-dominance and socioeconomic background. Together, these findings illustrate fundamental shifts taking place in the social and religious landscape and point to growing generational divides among the U.S. Latino population.

## **LGBT in Changing Times: Attitudes, Experiences and Growing Acceptance**

### **Political Party, Gender and Permissiveness: Understanding Attitudes Towards Same-Sex Marriage**

S. Kelley, *Yale University*

C. Kelley, *Yale University*

We analyze attitudes towards same-sex marriage, particularly the reciprocal effects of political party identification. Structural Equation Modelling analysis of a large national sample of the US (N=2295) suggests that Democratic Party identification makes both men and women more supportive of same-sex marriage; but conversely strong support for same-sex marriage makes men 10% more likely to choose the Democratic Party. Analysis of ISSP data shows that political party was not important in the US, Western Europe, or Eastern Europe fifteen years ago, suggesting a unique politicization of the debate in the contemporary US. Furthermore, there are significant gender differences in attitude formation: for women heterosexual permissiveness is twice as important as it is for men in creating support for homosexuality.

## **Pew Studies of the LGBT Population and Attitudes toward Them in 39 Countries**

Scott Keeter, *Pew Research Center*

This presentation will feature insights from the Pew Research Center survey of the LGBT population, based on a survey of 1,197 self-identified lesbian, gay, bisexual and transgender adults 18 years of age or older, interviewed online. Topics will include perceptions of the LGBT population regarding public acceptance and discrimination toward the community, as well as the priorities for policy changes and opinions on specific issues. Differences in attitudes about homosexuality among nations, and how these have changed in recent years, will also be discussed. In addition, the presentation will describe methodological issues involved in surveying this population.

## **The Political Consequence of LGBT Identity and Consciousness**

Ken Sherrill, *Hunter College, CUNY*

Andrew Flores, *University of California, Riverside*

For over a generation, students of public opinion and voting behavior have observed the emergence of LGBT voters as a distinctive political group. Unlike most other politically distinctive groups, commonly rooted in racial, religious, and ethnic households, LGBT identity is not transmitted generationally within the family. Rather, it is most reasonable to assume that LGBT people are randomly distributed about the population at birth and are, by and large, born into heterosexual, gender-conforming households. Earlier research finds that LGBT Americans tend not to become aware of their same-sex attraction until they are teenagers and do not begin to tell others about their sexual orientation until several years after that. Egan, Edelman, and Sherrill (2008) find that it is at this developmental stage that LGBT people change their political orientations to be more Democratic and liberal. Through such identity formation processes, LGBT people emerge as a cohesive and distinctive political group.

Our paper analyzes two waves of on-line polling conducted by Harris Interactive prior to the 2012 presidential election, with one wave in August of 1,190 LGBT respondents and 1,367 respondents from the public at-large, and with the second wave of about 500 respondents of each group conducted at the end of October, prior to the general election. In particular, we look at levels of consciousness and identity strength among LGBT respondents and how they affect partisanship, issue priorities, presidential vote, and levels of political engagement. In keeping with the general themes of this panel, we also compare the 2012 data with a 2003 Harris Interactive poll of LGBT respondents, most notably to see how the political priorities of the gay community have shifted in the course of a decade in which their rights have been the subject of increasingly partisan and contentious debate.

## **The Change in Attitudes toward LGBT through Surveys and in Personal Experience**

Murray Edelman, *Edelman Research*

The trends in attitudes over the past 40 years will be illustrated with some of Murray Edelman's personal experience. The trends in the LGBT population from the exit poll will also be discussed.

Murray is the first openly gay president of AAPOR and has also been recognized for his activist work at the start of the gay liberation movement. For 10 years, he was the Editorial Director of Voter News Service, the exit poll media consortium.

## **Geographic Information Systems (GIS) Methods and Tools for Survey Research**

### **Using GIS-Based Modeling to Understand the Potential for Response Bias in a Web Survey**

Ned English, *NORC at the University of Chicago*

Lee Fiorio, *NORC at the University of Chicago*

Michael Stern, *NORC at the University of Chicago*

Web surveys have undergone considerable study in recent years as an alternative to other modes, due to their potential for cost and speed advantages without the same coverage drawbacks associated with random-digit dial telephone surveys. One challenge has been how to recruit a household to a web survey using a probability-based mechanism, with physical mailings to selected address being one option. Our paper presents results from two studies where households were contacted both via physical mail and emails that were matched to selected addresses in a diverse state with rural, urban, and suburban environments. We use spatial modeling within geographic information systems (GIS) to understand how the kinds of people who respond to web surveys at as recruited by either mode of contact may differ from the population at-large. In so doing we consider not only what categories of households respond to a given mode or contact method, but also what kinds of places might be subject to a coverage bias by either contact method. Examples of methods employed and discussed include Moran's I, Spatial Lag Models, and LISA Statistics. Our research is useful for survey methodologists who are considering implementing a web survey and want to understand coverage limitations.

### **Geoscreening: Effective Use of Locator Services in Survey Recruitment**

Timothy Michalowski, *Abt SRBI*

Dara Seidl, *Abt SRBI*

Successful representation of marginalized groups in survey research often requires weighted sampling in unusual boundaries, such as distinct neighborhoods. One of the challenges in recruiting participants for telephone surveys is ensuring home addresses are located in the necessary target study areas before the full survey is conducted. Over-recruitment of participants can occur when home addresses cannot be verified as inside or outside the specific study areas until after the participant completes the full interview, and a final home address is verified typically from incentive mailing addresses. The ability to determine instantaneously over the phone whether a potential participant qualifies geographically for a study is a development that maximizes efficiency and prevents cost overruns. In a 2012 survey on the "Stop and Frisk" policy in New York City, Abt SRBI first implemented real-time "geoscreening" technology to determine whether the potential recruit qualified or screened out based on geography. This presentation reviews the technology behind geoscreening, the resulting time and budgetary savings, solutions for large-scale implementation, and how geoscreening is currently being applied in other survey projects throughout the country. With the rapid increase in cell-phone-only populations, the use of real-time GeoScreening offers a method to target cell-phone-only populations in any customized study area.

## **Coverage Comparison of Various Methods of Using the Postal Frame for Face to Face Surveys**

Stephanie Eckman, *Department of Sociology, University of Mannheim*

Various methods of sampling addresses for in-person surveys from the United States Postal Service Address list have been proposed (see, for example, Dohrmann et al 2012, Lepkowski 2010, Harter et al 2010). The methods have different impacts on the costs of frame construction and the proportion of the population that is covered. They also result in different design effects due to clustering. This paper will discuss the different impacts of the methods.

## **The Use and Limitations of Ground Captured GPS Coordinates for Address-Based Samples and In-Person Surveys**

Joseph McMichael, *RTI International*

Jamie Ridenhour, *RTI International*

Michael Keating, *RTI International*

Karol Krotki, *RTI International*

In-person surveys can gain valuable information through the GPS chipsets that are now commonly available on most smartphones and tablets. GPS coordinates captured during data collection are a useful tool for ensuring data quality by detecting interviewer error and fraud in addition to measuring error associated with geocoding addresses for address-based sampling (ABS) frames. We analyze results from ground-captured coordinates from Evaluation of Public Education Campaign on Teen Tobacco, an in-person nationally representative household ABS survey. We explore data quality and geocoding error by comparing the recorded location of the survey to the geocoded location of the housing unit; if the distance between locations is too great data quality could be compromised. We provide distributions displaying the distance between ground truth and geocoded location, examining associations between the degree of error and other attributes of a segment such as address type, urbanicity, and strength of the GPS signal. We suggest strategies for survey researchers to separate the wheat from the chaff – identifying suspicious coordinates without introducing too many false positives. Finally, we use this data to provide baseline measures comparing the geocoding accuracy of three address geocoding vendors.

## **Putting GIS on the Survey Research Map: Exploring Geoinformatics Principles for Enhancing Sample Survey Design and Analysis**

Trent D. Buskirk, *Marketing Systems Group (MSG)*

Dennis Dalbey, *Marketing Systems Group (MSG)*

Nathaniel Bordy, *Marketing Systems Group (MSG)*

John Zabrenski, *Marketing Systems Group (MSG)*

As geographic information systems (GIS) continue to mature as a distinct scientific field, applications of geoinformatics are emerging in many research settings. In particular, technological advances in mapping and remote sensing have allowed researchers to better understand public health outbreaks, consumer buying patterns, the distribution of voters by party across the U.S., the travel patterns of smartphone owners via passively collected geo-coordinates, among others. Survey researchers have begun to explore the benefits of mapping for both design and analysis including application of density maps to create sampling strata or thematic maps to provide a bird's eye view of survey response. In this paper we will explore how the basic principles of GIS can be applied to enhance sample survey designs and analysis. In particular, we will demonstrate the application of GIS techniques to five aspects of survey

design, analysis and operations including: (1) how to construct sampling frames using geo-coding; (2) how to use geo-spatial smoothing techniques to enhance and identify targeted strata; (3) how to use geo-mapping techniques to create optimal area probability designs that are representative of a chosen geography; (4) how to apply geographic mapping and geo-spatial measurements to optimize the fielding of in-person interviewing and (5) how to apply geo-distance measures and mapping to inform and enhance survey nonresponse adjustments. This presentation will also include a brief overview of some of the key terms from GIS that should be on every survey researcher's radar.

## **Old Paradigms/New Applications**

### **Exact Replication of Question Design Experiments from Schuman & Presser**

Henning Silber, *Stanford University*

Jon A. Krosnick, *Stanford University*

Tobias H. Stark, *Stanford University*

Annelies G. Blom, *University of Mannheim*

Numerous experiments on questionnaire design were conducted more than three decades ago in the U.S. (e.g., Cantril 1940; Gallup 1941; Schuman and Presser 1981). And many of the current best-practice standards in the field are based on the findings of these studies. Some especially important wording and format effects were particularly strong among people with less formal education (Narayan and Krosnick 1996) who have limited cognitive skills (Ceci 1991) and therefore tend to satisfice, which seems to have caused these response effects (e.g., Krosnick 2002; Narayan and Krosnick 1996; Knauper 1999). Because the American public has become increasingly educated in recent years, the magnitudes of the effects observed decades ago may be smaller now if rising education means increasing cognitive skills.

In this study, we test whether the old findings of response effects are still valid today. To this end, we are replicating thirteen experiments on response order, acquiescence, question wording, no opinion response options, question order, and question balance from Schuman and Presser's (1981) classic book. Our data will be collected from a national probability sample of the general American population ( $n = 4,000$ ) with email addresses. The survey will be conducted in December 2013 with the Gallup Panel. For each experiment, we will replicate Schuman and Presser's analyses and test whether the distribution of answers differs significantly across question order/formats. Second, we will test for differential effects in education subgroups and subject all experiments to a meta-analysis to test whether education moderates the strength of the response effects. If the older results are not replicated, this will suggest that the old findings and guidelines for question design should perhaps be considered dated and that new recommendations for optimal questionnaire design should be developed.

### **Nonresponse Error in a Total Survey Error Context**

Cong Ye, *American Institutes for Research*

Roger Tourangeau, *Westat*

Survey researchers advocate survey designs that minimize the total survey error for a given cost (see e.g., Biemer and Lyberg 2003; Groves 1989). However, studies are often forced to focus on one or two of the many sources of error, usually the sampling variance and perhaps one or two potential biases. Furthermore, the relative importance of the errors that are studied is rarely evaluated; this means researchers have no guidelines for resource allocation. To our



knowledge, the four major types of errors—coverage error, sampling error, nonresponse error, and measurement error—have not been investigated together within one study; instead, individual errors have typically been the focus.

Taking advantage of records data for all sample members in two surveys (one on voting and the other on academic performance), we evaluated the importance of nonresponse error in a total survey error context. We study three sources of error in one survey and four sources of error in another survey and evaluate their relative importance. In addition, we assess the amount of reduction in error that postsurvey adjustments can achieve. In both data sets, measurement biases tended to swamp all other forms of error for the undesirable characteristics and nonresponse tended to produce the next largest error. Weighting adjustments generally help reduce the biases, but reduce the total error by no more than 10%. We also examined estimates for two demographic variables (sex and age). For these variables, the overall errors are much smaller and measurement bias is no longer the main source of error.

### **Experiment Designs in Population-Based Survey: How Much Have We Learned About Between- and Within-Subjects Designs?**

Bo MacInnis, *Stanford University*

Jon A. Krosnick, *Stanford University*

Scholars from public opinion, political science, political psychology and other social sciences are increasingly using nationally representative surveys to conduct population-based experiments (Mutz, 2011). In designing an experiment to determine the effect of an independent variable (experimental condition) on some dependent variable, an investigator faces the choice of manipulating the independent variable between or within subjects. In a between-subjects design, respondents are randomly assigned to receive one experimental condition; in a within-subject design, each subject receives more than one experimental condition. The assumption seems to be made in many studies that the independent variable will have the same effect on the dependent variable with the administration of either method, consequently favoring within-subjects design to gain greater statistical precision. Many experiments have shown this assumption fails to hold in many situations, and within- and between-subjects experimental designs yield disparate results with treatment effects obtained through the former are biased and illusory.

In this study, we review the literature on between- and within-subjects experiment designs embedded in surveys from various domains. We then illustrate the disparity in results through between- and within-subjects experimental designs with new evidence from studies on public monetary valuation of producing a new good or service, or restoring a damaged good or service. A number of studies have assessed public monetary valuations of goods, services, or policies, e.g., harm caused by the Exxon Valdez oil spill (Carson et al, 1992), valuations of health care programs (Diener, O'Brien and Gafni, 1998), monetary assessment of cultural goods in arts and culture (Noonan, 2002). Using various survey experiments we conducted on assessing public willingness to pay for a reduction in future global warming employing both within- and between-subjects designs, we illustrate the merits and perils of two experimental designs. Implications for general survey experiment designs are discussed.

## **Is Social Exchange Theory Still Relevant for Explaining Why People Respond to Surveys?**

Don A. Dillman, *Washington State University*

In the 1970's I applied social exchange theory, with its joint emphasis on reducing costs, increasing benefits, and creating trust, to creating a framework for how to improve response to surveys (Dillman, 1978). The social exchange framework, originally developed as a way of explaining and predicting interaction with other people, has been used by survey designers as a frame of reference for designing data collection protocols for more than three decades. In this paper I address the question of whether social exchange remains useful as an explanation for why people do or do not respond to survey questionnaires in today's rapid-fire, multi-media social environment. To do this, I first describe the initial development of this survey design framework, its strengths and limitations as I saw them at the time, and the reasons for its wide-spread use. Next, I will discuss whether social exchange concepts are adequate for explaining survey response in our increasingly mixed-mode survey environment, and how this theoretical perspective compares and contrasts with other theoretical possibilities. My goal is to suggest the need for gaining a better understanding of the barriers to survey response and how to overcome them.

## **A New Paradigm for Survey Methodology: Replacing Static Quality Concepts by a Modern Process Quality Perspective**

Peter Ph. Mohler, *University of Mannheim Germany*

On July 29 2013 a post to AAPORnet titled "being proud of survey research" ended with a bold statement: "Thus it seems timely to turn our attention from searching for errors to optimizing survey process quality and at long last defining benchmarks for good surveys that are fit for their intended purpose. The concepts and tools are already there, waiting to be used to our benefit." The proposed paper will expand and elaborate this statment. In short the new framework of survey research will be process quality oriented, always on the learning side and, in the end, provide a scientifically and engineered sound base why people can be proud to be survey researchers. This will be achieved by a shift of perspective as it is already under way in survey methodology: from static quality indicators such as response rate or response bias to dynamic, production process related aspects. Examples are using Paradata as process quality indicators or balancing quality for all stages of the Survey Production Process (Survey Life Cycle). This new perspective focusing on the complete Survey Life Cycle inevitably results in a replacement of Error as the core concept by Survey Process Quality as the overarching concept. In addition this also means increasing emphasis on process engineering and less on sample statistics (without diminishing its important role). In concentrating on Error we might have changed the question we always wanted to answer from "can I make valid and reliable inferences from my sample survey on the whole populations" to a much easier to answer question such as "is there nonresponse bias and/or measurement bias". And as we learned in the past decade, to answer the latter is probably not good enough. The new paradigm will force us to answer the tough questions.

## **Reviewing and Revising Pre-Election Polling Methods: Insights From the Gallup Election Review**

### **Likely Voter Modeling in Pre-Election Polls**

Michael W. Traugott, *University of Michigan*

This paper will analyze different models for estimating likelihood to vote (LV), based upon different questions to form a LV index as well as both cutoff and logistic regression models to construct weights for individual respondents. They will be evaluated in pre-election surveys in both states as well as in comparison to differences in pre-election and post-election self-reports.

### **Question Wording Experiments to Minimize Over Reporting of Expected Vote**

Frank M. Newport, *Gallup, Inc.*

Alternative question wordings were randomly administered and used to measure pre-election and post-election self-reports of voting in the 2013 gubernatorial elections. Using a source monitoring approach effectively reduced over-reports compared to a standard questions, suggesting an improved way to assess likelihood of voting in pre-election surveys.

### **Mode Differences in Pre-Election Poll Estimates of Voting**

Michael W. Traugott, *University of Michigan*

As part of the experiments conducted in conjunction with the elections in New Jersey and Virginia, Gallup used RDD telephone samples as well as a web survey to assess likelihood of voting and candidate preferences. This paper will highlight observed difference between the two study designs and evaluate their suitability in terms of resulting sample characteristics and their consequences for likely voter models and candidate preference distributions.

### **A Comparison of Pre-Election and Post-Election Self-Reports of Voting**

Frank M. Newport, *Gallup, Inc.*

This paper addresses the issue of quality of pre-election self-reports of intent to vote and post-election reports of actual voting. The paper will focus on over and underestimation of both turnout and its consequences for candidate preference with a particular focus on specific subgroups in the electorate.

## **Paradata Applications in Production: The Challenges of Change**

### **Incorporating Paradata Metrics into Daily Survey Management**

Barbara C. O'Hare, *U.S. Census Bureau*

Tamara S. Adams, *U.S. Census Bureau*

John A. Wilen, *U.S. Census Bureau*

The implementation of adaptive survey design presents challenges as field data collection processes are modified to achieve strategic survey goals. This paper discusses progress in implementing survey metrics and dashboards for use by survey managers, including survey directors and field staff managers across six demographic household surveys conducted by the Census Bureau.

As survey analysts and survey operations staff collaborated in the past year at Census, we have confronted the challenges of putting paradata into practice in a systematic, strategic way. This paper provides an update on the Unified Tracking System of survey paradata integrated across multiple operations systems and providing daily reporting to a wide range of end users. The system includes a first generation daily dashboard of key survey performance metrics at the national and regional office levels. We will describe the process to design the system to include survey progress, quality and cost metrics. In addition, we will note and discuss the challenges in getting survey managers to embrace the system and to recognize how to use the paradata reports.

The survey analysts and the survey operations staff are working together to develop new ways toward the goal of addressing survey goals related to cost and quality. We will share some of the ways that the daily reports are being used by field staff managers in our regional offices to identify interviewers and interviewing teams that are performing above or below expectation, leading to intervention. We will also discuss the challenges of integrating higher level survey metrics into daily monitoring and managing toward specific outcomes, such as cost per case targets.

### **Use of Response Propensity Scores to Direct CAPI Field Activity**

Jamey Christy, *U.S. Census Bureau*

The declines in survey cooperation and increasing budget constraints are challenging all survey organizations, including the Federal Government. In this environment, survey organizations are taking steps to better address these issues during the data collection period. One way this can be done is by evaluating cases on their response propensity and making management decisions on the allocation of resources to the cases. In this paper, we discuss how the results of a response propensity model based on contact history is implemented in the American Housing Survey CAPI production environment, and how the propensity information was used in case management decisions by the staff overseeing the data collection activities. During the past few years, the Census Bureau has introduced propensity models and their potential use in some of the demographic household surveys the Bureau conducts. Statistical analysis of historical data indicates that we can estimate relative propensity fairly well.

Discussions with operations staff have furthered their understanding of how a propensity indicator could be helpful in case management, but has also raised concerns about how use of this information could result in unintended consequences by focusing on particular cases. We will report our experience in estimating propensities in a production environment from both the statistical analyst and the field manager's perspectives. We will present the challenge of building the model based on limited historical data, the process of nightly estimations of propensities, the mechanism for providing this information to the data collection staff, and the use of these data in field activities.

### **Improving Telephone Survey Data Quality with Real-Time Access to Paradata**

Andy Weiss, *Abt SRBI*

Call center interviewers can dial over thirty records per hour. The dialing process generates a huge volume of paradata. The pace of telephone interviewing requires close interviewer management to avoid problems impacting a substantial amount of sample.

We describe ways to classify and aggregate this data to guide sample and interviewer

management and address the dangers and limitations of this data. We demonstrate an approach to this volume of data that enables better mid-shift decision making using a set of filtering tools that is applicable across a wide range of projects. We will present a manner of thinking about this data to improve call center efficiency, survey response and general data quality. In particular, we have built automated reports that can be run on demand to guide monitoring, re-training and other interviewer feedback efforts. We will also discuss the risks this data poses to interviewer morale and call center culture. Our set of customizable reports is built around a sensitivity to the difficulties telephone interviewers face with increasingly hostile potential respondents.

More broadly, we will discuss uses of call center paradata to trigger more qualitative interaction with specifically targeted individual interviewers.

### **Challenges of Implementing Adaptive Design Strategies in Production Settings for Large Records in an Establishment Survey**

Melissa Mitchell, *USDA National Agricultural Statistics Service*

Kathy Ott, *USDA National Agricultural Statistics Service*

Jaki McCarthy, *USDA National Agricultural Statistics Service*

With increasing nonresponse rates, biased survey estimates are a growing concern. Establishment surveys conducted by the USDA's National Agricultural Statistics Service (NASS) are no exception. The annual Agricultural Resource Management Survey (ARMS) conducted by NASS provides important economic data on US farms, but has relatively lower response rates than many other NASS surveys. In order to target nonrespondents, NASS began to use decision tree models to identify likely nonrespondents using not only survey data but auxiliary data as well. Using Census of Agriculture data, NASS has identified characteristics associated with nonresponse in ARMS and thus identified subgroups of establishments that are less likely to respond (Earp & McCarthy, 2009, 2010). These subgroups are identified using an ensemble of decision tree models.

In 2012, we used these models, coupled with impact scores which determine how important the operation is to the estimate, for our targeted data collection efforts. Using the models with the impact score allowed us to target data collection efforts not only on likely nonrespondents, but on those nonrespondents with the most impact on the economic estimates. The specialized data collection efforts we used were face-to-face contact, using interviewers or regional office employees with the best past response rates for harder cases, and providing enumerator incentives for the hardest cases. We had the enumerators fill out a follow up form for each operation they visited to see what techniques they used, how well the techniques worked, and if they were following our instructions. We found that there was reluctance to follow our instructions, especially when it came to our high impact operations. This paper illustrates the challenges of doing an experiment in a production setting overall then focuses on the additional challenge using high impact operations in any experimental framework.

### **Using Paradata to Identify and Remove Vacant Housing Units from In-Person or Telephone Follow-Up Interviews**

Geoffrey Jackson, *U.S. Census Bureau*

Due to the decrease in survey self response and the cost associated with following up with nonresponding housing units, the Census Bureau is researching ways to use an adaptive design to maximize the efficiency of their programs. This research looks at identifying vacant

housing units while minimizing the error rates for this determination prior to the start of an in-person or telephone follow-up interview to reduce interview workload. One method to pre-identify vacant housing units is to use data collected by the United States Postal Service. In the 2010 Census, multiple pieces of mail were sent to a housing unit. When the United States Postal Service was unable to deliver mail to an address, they identified that piece of mail as undeliverable-as-addressed. A vacant living quarters is one of the twenty reasons that a piece of mail is identified as undeliverable-as-addressed. The resulting outcome of each piece of mail sent out for a survey provides insight on the occupancy status of that address; this includes the advance letter, reminder postcard and questionnaires. This research looked specifically at housing units with undeliverable-as-addressed vacant reason codes from multiple pieces of mail associated with the 2010 Census mail out. The multiple confirmations of vacancy from the United States Postal Service increase the likelihood of the housing unit actually being vacant. To obtain better results for predicting the vacancy of specific housing units, the undeliverable-as-addressed codes were paired with additional data collected by the United States Postal Service and data about building structure from the address frame. Final results from the 2010 Census were used to compare against the vacancy status that would be assigned based on UAA information and to evaluate potential impacts on hard-to-count populations by not following up with these housing units.

## **Public Opinion: Theory, Effects and Measurement**

### **Two Sides to Every Story: What Happens when Public Opinion and Expert Opinion Conflict?**

David L. Vannette, *Stanford University*

Sean J. Westwood, *Stanford University*

Public opinion data are increasingly central to news media coverage of stories ranging from the campaign horserace and support for government policies to celebrity red-carpet fashion and new consumer products. This has led to members of the public increasingly being asked to share their opinions about a wide range of subjects, and these opinions are then being presented back to them by the news media at a greater rate than ever before. This trend has not gone unnoticed by researchers and a growing body of work is exploring the potential effects of public opinion on individual attitude formation and behavior. But little work to date has examined the relative influence of polls and other types of information. In this paper we experimentally test this question: when expert opinion and public opinion data are presented together which has a greater effect on attitude formation? We report the results of two studies, one utilizing a nationally representative Internet sample and the other using a convenience Internet sample. Using experimentally manipulated news stories, presented in the context of confirmatory or contradictory expert opinion, we examine the effects of published public opinion data on public support for politicians and policies. The results from these studies confirm that poll information does indeed act to shape attitudes. However, these studies extend this finding to provide novel evidence that these effects of public opinion data overwhelm other sources of conflicting information in a news article. These results have implications for the study of public opinion and the psychological effects of public opinion on attitude formation.

## **Mini-Publics and Public Opinion: A Survey-Based Experiment**

Shelley Boulianne, *Grant MacEwan University*

Government officials host many public consultation initiatives around diverse topics. These initiatives demonstrate respect for the democratic process as well as may increase the perceived legitimacy of the democratic system and its outcomes. In more intense forms of public consultations, a select group of citizens is given a large amount of information, then asked to deliberate on policy directions and make recommendations. The great concern with these “mini-publics” is whether they are representative of the broader public, public opinion, and whether under similar conditions, another group of citizens would reach the same policy recommendations.

This paper will present the results of survey-based experiment that assess the impact of a “mini-public” on the public opinion of an RDD sample of the general public (n=400). Approximately half of respondents were asked their opinion on a number of policy directions to address climate change and energy issues. The other half of the respondents were given the same list of policy directions, but were informed that these policy directions are based on recommendations by a randomly selected, diverse group of 60 citizens who met for six Saturdays to examine climate and energy issues and provide recommendations (“mini-public”). The two conditions are compared for varying levels of support for the numerous policy directions, as well as varying levels of trust in government and political efficacy. While there were no large differences in support for different policy directions, there were substantial differences in political trust and efficacy for the two conditions. Respondents who were informed about the “mini-public” reported higher levels of trust and efficacy, compared to the condition where respondents were not informed about the “mini-public”. The findings suggest that these mini-publics help generate legitimacy in the political system, not through policy legitimation, but through building trust and political efficacy among citizens.

## **Potential Effects of Government Shut Down on the Federal Statistical System**

Jennifer Hunter Childs, *U.S. Census Bureau*

Ryan King, *U.S. Census Bureau*

The Federal Statistical System Public Opinion Survey has been gathering data on public understanding of and trust in Federal Statistical Agencies and federal statistics over the course of about two years. These public opinion data are meant to enable federal statistical agencies to better understand public perceptions, which will provide guidance for communicating with the public and for future planning of data collection that reflects a good understanding of public perceptions and concerns. From February 2012 through September 2013, the Census Bureau added 25 questions nightly onto an ongoing data collection by the Gallup Daily Tracking Survey. From October 2013 through March 2014, the Census Bureau will add 10 questions to the survey.

During the government shut down in October 2013, we reinstated a few previously fielded items that focused on cost savings as a reason to use administrative records to examine any possible effect on trust of a government shut down. Initial results show that trust in federal statistics decreased during the time of the shutdown. Perceived relevance of federal statistics also slightly decreased and feelings that statistical agencies invade people’s privacy also slightly increased. Fortunately, there was no change in the item that measures whether people believe that federal statistical agencies keep their information confidential. The paper will discuss more in-depth results as well as limitations and implications for the federal statistical system.

## **Debunking the Myth of American Isolationism**

Dina Smeltz, *The Chicago Council on Global Affairs*

Greg Holyk, *Langer Research*

Earlier this year, US public opposition to military action against Syria became the evidence for many commentators of a new isolationist mood in America. But 40 years of Chicago Council Survey data show that isolation has never been a prevailing tendency among the American public in contemporary times.

Although the world looks much different today than it did in 1974, when the first Chicago Council Survey (previously CCFR) was fielded, there are some striking similarities between the 1974 and most recent 2012 findings. Americans have become more selective in their preferences for international engagements for some of the same reasons they were selective in the 70s, because of constraints on US economic resources, war weariness and a desire to scale back from unpopular foreign entanglements. But that doesn't mean the public is turning inward.

Over the past 40 years – spanning the eras following the Vietnam war, the fall of the Berlin Wall, and the September 11th attacks – about two-thirds of Americans have consistently preferred that the US take an active part in world affairs. Solid majorities have agreed on the most important foreign policy goals. Americans tend to support maintaining our defense budget. And majorities have also consistently expressed support for maintaining our commitments to NATO. Taken together, these data will show that Americans want to participate in world affairs, but in a less interventionist way.

The preferred nature of such engagement and its priorities also has shifted over time based on changing domestic and international realities. In this paper we will show how many of the mistaken examples of isolationism are really just variation in the way Americans would prefer to engage internationally.



**Friday, May 16**  
**1:45 p.m. – 3:15 p.m.**  
**Concurrent Session E**

## **Cross-Cultural Survey Research**

### **Which Language Should We Use for Interview in Multi-Linguistic Environment – Results of Comparative Experiments**

Anna Andreenkiva, *CESSI (Institute for Comparative Social Research)*

Language of interview is not always obvious choice in many cultures, nations and political contexts. Researchers are confronted with such choice in regions with multi-lingual population which is growing in number in many regions because of globalization of education, migration, multi-ethnic marriages and other factors. The decision about the language of interview is taken on project planning stage and on the field stage of the survey. On the planning stage several types of considerations can influence the choice of languages for interviews – linguistic, organizational (costs of the translation, visibility of high quality translations, interviewers with particular language skills, etc.), social and political. On field stage the choice of language depends on the set of languages pre-defined by researcher, cultural norms and traditions in interaction, functional image of the language among target group, linguistic skills of actors, type of interaction between interviewer and respondent, direct influence of interviewer on respondent, personal preferences.

The paper describes the current practices and methodological experiments on the choice of language and its consequences conducted in multiple public opinion polls in countries of the former Soviet Union conducted by CESSI in recent years. Languages fulfill different functions as mean of communication but also as symbolic sign of belonging to particular groups and sharing particular views - which can substantially influence the results of the survey. Paper shows that in multi-lingual situation languages different languages are used by respondents for fulfilling different tasks (inter-family communication, outside communication, professional communication, media consumption). Languages are also not equal in the linguistic skills (oral everyday vocabulary, oral social vocabulary, writing and reading skills). All these should be taken into account when taking decision which language should be chosen for particular mode of interviewing (self-completion, face-to-face interview, telephone survey etc.), topic of interview etc.

### **Assessing Translations: How do Backtranslation and Committee Approach Compare**

Alisu Schoua-Glusberg, *Research Support Services*  
Ana Villar, *City University*

This paper empirically compares the efficacy of two methods: backtranslation and committee approach. Twenty years ago, the most commonly used method for assessing questionnaire translations was backtranslation, where a translated question was translated back into the original language and this new text compared to the original question. The rationale was that this comparison allowed researchers to assess the quality of the translated text and identify translation problems without knowing the target language.

Many still believe backtranslation is the industry standard, despite federal agencies and university researchers increasingly adopting team or committee approaches. Committee approaches involve two or more translators working on a translation that is then discussed at a meeting where translators and other experts come together.

Previous research has compared backtranslation to individual expert reviews (Harkness et al, 2009), but no studies have compared backtranslation to the committee approach. This paper intends to fill this gap. 66 items were translated from English into Polish by three translators independently. In the committee condition, the three translations were discussed at a meeting where a final version was agreed. In the backtranslation condition, the Polish items produced by the translator with the strongest credentials were sent to a backtranslator, and the two English versions (original and backtranslated) were compared by non-Polish speaking survey experts.

In this presentation we discuss how each approach performed in capturing translation mistakes and other problems and we present next steps to continue this investigation.

### **Utility of Nonverbal Behavior Coding for Detecting Comprehension Difficulties Across Race/Ethnic Groups**

Timothy P. Johnson, *University of Illinois at Chicago, Survey Research Laboratory*

Allyson Holbrook, *University of Illinois at Urbana-Champaign*

Sharon Shavitt, *University of Illinois at Chicago, Survey Research Laboratory*

Marina Stavrakantonaki, *University of Illinois at Chicago, Survey Research Laboratory*

David Sterrett, *University of Illinois at Chicago, Survey Research Laboratory*

Young I. Cho, *University of Wisconsin-Milwaukee*

Noel Chavez, *University of Illinois at Chicago*

There is little evidence available regarding whether or not the systematic coding of nonverbal respondent behaviors can be used to provide insights into respondent processing of survey questions. In this paper, we examine patterns of nonverbal behaviors in a sample of 400 interviews with African American, Korean American, Mexican American and non-Hispanic white adult respondents interviewed in Chicago. In addition to describing the process used to systematically code nonverbal behaviors, we will also investigate: (1) the degree to which these nonverbal behaviors provide information consistent with (or in addition to) established sets of verbal behavior codes that have been previously demonstrated to be useful for identifying comprehension and mapping problems; and (2) the degree to which these associations are consistently replicated across race/ethnic groups. We will conclude with recommendations regarding the usefulness and comparative advantage of employing nonverbal behavior codes as measures of cognitive processing difficulties, as well as their cross-cultural utility.

## **Measuring Culture Across Racial and Ethnic Groups in the U.S.: Exploring Challenges and Solutions**

Allyson L. Holbrook, *Survey Research Laboratory, University of Illinois at Chicago*

David Sterrett, *University of Illinois at Chicago, Survey Research Laboratory*

Marina Stavrakantonaki, *University of Illinois at Chicago, Survey Research Laboratory*

Timothy P. Johnson, *University of Illinois at Chicago, Survey Research Laboratory*

Sharon Shavitt, *University of Illinois at Chicago, Survey Research Laboratory*

Young I. Cho, *University of Wisconsin-Milwaukee*

Noel Chavez, *University of Illinois at Chicago*

Saul Weiner, *University of Illinois at Chicago*

Culture has been defined as “the collective programming of the mind which distinguishes the members of one group from another” (Hofstede, 1980, pp. 21-23). Measures of culture such as Triandis’ cultural orientations scale and Schwarz’s cultural values scale have been used to assess cultural differences among residents of different countries and among members of different racial and ethnic groups in the U.S. Measuring culture across countries or racial and ethnic groups relies on the assumption that the measures assess the same underlying constructs within each group. One reason for concern about measures of culture is that these measures may be affected by response effects that vary across cultural groups (e.g., acquiescence response bias or nondifferentiation). Because of this concern, some researchers have recommended that responses to scales designed to measure culture be standardized within cultural group. However, as yet there is no consensus for how (or when) cultural orientation measures should be adjusted when comparing means across groups, and different studies employ different methods. Using data from two studies in which a total of nearly 1000 adult non-Hispanic Whites, African-Americans, Mexican-Americans, and Korean-Americans from the Chicago area completed the cultural orientation scale and the cultural values scale in a self-administered questionnaire, we examine the associations among cultural orientation subscales as well as the factor structure of these measures across and between racial and ethnic groups to assess cross-group equivalence. Our analyses focus on the effect of different forms of standardization (e.g., within participant, within cultural group) on the observed results. We also compare our findings to those observed in numerous past studies. Finally, we discuss the implications of our findings for assessing culture across racial and ethnic groups in the U.S.

Hofstede, G. 1980. *Culture’s Consequences: International Differences in Work-related Values*. London: Sage Publications.

## **Why Do We Need Translated Questionnaires? Insights from a Large-Scale Study of Muslims and Christians.**

Inna Becher, *Institute for Employment Research (IAB)*

Most general population and migrant surveys conduct interviews in a single language only. Translations of the master questionnaires into foreign languages are not common due to additional costs of (back-) translation or of pretests that have to be conducted in other study languages.

But which bias do we introduce to the data if we systematically exclude those subgroups without knowledge of the national language? And how does the exclusion of these subgroups affect the substantive results of the survey? We investigate these research questions on the basis of a study of Muslims and Christians with migration background in Germany. For this purpose we asked respondents about their reasons for choosing a translated version. Using multivariate

analysis, we examine the differences between the two groups – with and without sufficient knowledge of German – concerning demographic characteristics and substantive survey variables. We especially analyze differences on attitude questions about gender roles, but also on religiosity and religious affiliation of the respondents.

Our analysis shows that the selection process based on the language issues has a considerable biasing effect. The group of ‘potential non-respondents’ (persons without knowledge of German) is dominated by first-generation migrants, Muslims and persons with children in the household. Furthermore, the integration and employment status play a significant role in the selection process. For this reason, we strongly recommend the use of translated questionnaires if the migrant population is under study.

## **Cell Phone Survey Sampling**

### **Exploring the Impact of Various Control Total Sources for Adjusting for the Cell Only Population in the California Health Interview Survey**

Ismael Flores Cervantes, *Westat*

Matt Jans, *UCLA Center for Health Policy Research*

Trent D. Buskirk, *Marketing Systems Group*

Dual-frame (cell phone and landline) random digit dial (RDD) samples greatly reduce coverage errors in a variety of survey estimates compared to landline-only RDD samples. However, for state and local surveys, the choice of control total sources for post-stratification of the cell phone population has been limited. Most post-stratification adjustments implemented in dual-frame survey weightings require estimates of three phone user populations: landline-only, cellphone-only, and dual-users. However, sub-national estimates of these three populations are hard to find across consistent levels of geography and may have errors larger than is acceptable for a control total source. For the past 3 cycles, the California Health Interview Survey (CHIS) has used U. S. Census multi-state regions estimates from the National Health Interview Survey (NHIS) published by the National Centers for Health Statistics (NCHS). Two new sources of cell phone population estimates for smaller geographic areas have become available recently. The first source includes NCHS state-level estimates and model based county-level estimates. The second source is Marketing Systems Group (MSG) that produces quarterly county level estimates for cell phone-only households based on phone number assignments and administrative databases. Since these estimates are produced for smaller geographical areas, their use can potentially improve estimates from state and local dual-frame RDD surveys. However, these sources are not the same and differences in the available geographic levels, the frequency they are produced, and available phone use categories determine how these are used in the poststratification adjustment. This paper shows the impact of using these new sources during weighting on several health-related estimates correlated to telephone usage such as smoking status and alcohol use. These estimates are compared to the traditional region-level source used in CHIS. Qualities of the weights and logical pros and cons of each source are addressed.

## **Strategies for Increasing Efficiency of Cellular Telephone Samples**

William Robb, *ICF International*

Kurt Peters, *ICF International*

Joshua Brown, *ICF International*

Ashley Mark, *ICF International*

Naomi Freedner, *ICF International*

Cristine Delnevo, *Rutgers University*

Daniel A. Gundersen, *Rutgers University*

A methodological study of indicators appended to a cell phone sample was conducted as part of the 2013 National Youth Adult Health Survey (NYAHS) to determine whether their use could increase the efficiency of cell phone surveys. The NYAHS relies on a national random digit dial (RDD) sample of cell phone numbers to conduct interviews with young adults. Two variables developed by the sample provider were appended to drawn sample: A Cellular Working Identification Number Service (Cell-WINS) indicator which identifies active numbers (active rate = 63%); and a billing ZIP code (append rate = 46%). To test accuracy, all records (those with and without indicators) were dialed and self-reported ZIP code was collected. Results revealed Cell-WINS to be an accurate indicator of active phone status (True Positive Rate (TPR) = 96%; True Negative Rate (TNR) = 67%), but ZIP-append less so (TPR = 88%; TNR = 38%). Productivity was also higher for Cell-WINS active vs. inactive sample but did not differ for sample with vs. without an appended ZIP code. The accuracy of the Cell-WINS flag may make it tempting use only “active” cell phone sample when conducting RDD surveys; however, this research suggests that doing so may introduce coverage bias due to systematic differences between these populations. In particular, the population that would be excluded appears to be lower in SES. Finally, comparison of the appended ZIP codes revealed that the accuracy of the ZIP-append at the ZIP-code level is low (36% match-rate) but increases with broader geography (e.g., 88% match-rate within Census region). Practical implications of these results for sample design are discussed.

## **Cell Phone Surveying in the Middle East**

John Lee P. Holmes, *Qatar University Social and Economic Survey Research Institute*

Abdoulaye Diop, *Qatar University Social and Economic Survey Research Institute*

Kien Le, *Qatar University Social and Economic Survey Research Institute*

Elmoqiera Fadlallah Elsaye Elawad, *Qatar University Social and Economic Survey Research Institute*

Isam Mohamed Abdelhameed, *Qatar University Social and Economic Survey Research Institute*

Anis Miladi, *Qatar University Social and Economic Survey Research Institute*

In his recent book on two decades studying Arab public opinion, Shibley Telhami barely mentions telephone surveys. Apart from Turkey and Israel, academic surveys in the region are face to face. Reasons include fear of monitoring, suspicion of solicitation and absence of a telephone frame. Last year Qatar University’s Social and Economic Survey Research Institute (SESRI) leveraged its unique blend of academic and governmental contacts to pioneer phone surveys. Like many developing countries, the Middle East skipped from sparse landline coverage directly into the age of the wireless. Twenty years ago most non-business calls were made from “phone centers” and few people had household landlines. Today, not only is cellular coverage in places like Qatar almost universal, it is likely higher than in the West. Here wireless services are more reliable than landlines which are purchased primarily in tandem with internet services.

In this paper we will review SESRI's recent telephone interviewing in Qatar. We'll discuss practical issues related to days and hours of calling, gender and interviewer effects, mode differences and concerns about security's potential impact on coverage that Telhami mentions in *The World Through Arab Eyes* (2013, p. 10). Using data from a telephone pilot study, field verification calling, two client phone surveys and a new Qatar Quarterly survey, which is developing a consumer confidence index, the results show a reliable coverage of Qatar's sub-populations, nationals and expatriates. In addition, unlike face-to-face, we have found it is possible to have opposite gender interviewing by telephone. So in place of asking "what will we do about cell phones," a question that faced American public opinion researchers a decade ago, we ask instead: "what's the best time to call you on your cell phone?"

## **Comparing Landline and Cell Phone Samples in the Survey of Consumer Attitudes**

Li Jiang, *University of Michigan*

Charley Jiang, *University of Michigan*

James M. Lepkowski, *University of Michigan*

Richard Curtin, *University of Michigan*

Dan Zahs, *University of Michigan*

The University of Michigan's Survey of Consumer Attitudes added a significant number of cell phone numbers to its landline telephone sample in July 2012. This change was made in order to increase overall coverage as more and more adults are not reachable by landlines. The 18-34 year old age group was of particular concern as due to response rate and coverage issues; this group has been under-represented in un-weighted data. As expected, the dual-frame design added importantly to the number of 18-34 year old respondents in the final sample.

However, due to the nature of cell phone sample, overall interviewing costs increased substantially and there was a net reduction in the survey response rates because of low cell phone sample response rates. A pre-screening process for landline telephone numbers has long been used to improve efficiency, and we wish to find a comparable procedure for cell phones. A cell phone pre-screen process based on a billing activity index could eliminate one-quarter of the cell phone sample numbers. We examine the effect of this cell phone pre-screening process on cost, calling pattern, work efficiency, response rates, and sample demographic composition for the period July 2012 to August 2013. We show that the cell phone sample pre-screening has the potential to reduce costs, increase the proportion of productive calls earlier in the monthly calling cycle, increase calling efficiency, and increase response rates slightly, all while maintaining the demographic composition of the sample obtained without pre-screening.

We also examine the impact of pre-screening cell phone sample numbers in the dual-frame sample on the bias of survey estimates. The pre-screened cell phone sample performed as well as the non-screened call phone sample in regards to the ability to obtain completed interviews for the 18-34 year old age group.

## **Localized Cell Phone Samples Utilizing Billing Zip Code and Rate Center Information: A Hybrid Approach**

David Dutwin, *Social Science Research Solutions*

Alisha Simon, *Minnesota Department of Health*

Kathleen Call, *University of Minnesota*

Stefan Gildemeister, *Minnesota Department of Health*

Cell phone sampling has undergone a sea change in the past half-decade as researchers try to find ways to increase the efficacy and accuracy of geo-location for a device that is, essentially, mobile. This is critical to research given that policy analysts frequently need to produce estimates from telephone surveys for specific geographies, within the U.S., within a given state, and even within a specific metropolitan area. Yet geographic precision in knowing the location of a person owning a given cell phone telephone number is quite limited.

While “rate centers” have now been utilized for local sample selection for some time, the availability of billing zip codes is now only in its second year. Billing zips, despite their geographical precision, have one major hurdle in that only about 60 percent of all cell phones have such information, and cell phone owners without a billing zip are substantially different than those who have a billing zip on a range of demographics. On the other hand, the main challenge in utilizing rate centers for geo-location is that they are not particularly precise, with about half of all cell phone owners presently not living in their designated “home” rate center.

The 2013 Minnesota Health Access Survey utilized a hybrid approach, sampling by billing zip where available, and rate center in sample for which there was no billing zip code. The data from the project allows us to make claims with regard to the relative efficacy of cell phone sampling and stratification by rate center only, by bill zip alone, or a combined approach. Each design is contrasted against each other on measures of coverage, incidence, geographic precision, and design effect. From these data we recommend “best practices” with regard to designing cell phone frames in order to accurately target small geographies.

## **Voting Behavior and Political Participation**

### **The People in Your Neighborhood: How Political Minority Status Affects Political Participation**

J. Quin Monson, *Brigham Young University*

Chris Karpowitz, *Brigham Young University*

Lindsay Nielson, *University of California, San Diego*

Kelly D. Patterson, *Brigham Young University*

Steve Snell, *Princeton University*

How does the way a voter relates to her community affect her political behavior? Political behavior, especially turnout, can vary significantly by community, but we do not yet fully understand the specific factors that produce this variation. In this paper we explore the relationship between community norms of partisanship, feelings of belonging in a community, and likelihood of turnout. We theorize that voters who fall outside of their community’s political norm are likely to identify as outsiders and that this self-awareness has a demobilizing effect.

We use data from two modules of the Cooperative Congressional Election Study (CCES) to explore what factors make a person feel like an outsider and how this outsider status influences

their political behavior, including intention to vote and actual voting behavior. We find that subjective feelings of similarity or difference from a community are indeed grounded in objective truth: citizens who are distant from the median voter in their community are much more likely to say that they feel like an outsider, but outsider status is also a function of an individual's personality traits and political interest. We also show that people who report feeling like an outsider are generally less likely to intend to vote, to report having voted, and to actually turn out to vote; these findings are robust across a variety of model specifications and are statistically and substantively significant even when other control variables are included in the models.

### **Ideological Congruence in Times of Economic Crisis**

Ioannis Andreadis, *Aristotle University of Thessaloniki, University of Michigan*

Eva H. Önnudóttir, *CDSS/Mannheim University*

Viktor Orri Valgarðsson, *University of Iceland*

This paper examines ideological congruence voting in times of a long-term economic crisis. We examine whether the economic crisis can make people vote for a party that has run on electoral populist promises about how to deal with the economy, without paying attention to large differences they have with it on other important issues. For this purpose we use the Greek 2012 and Icelandic 2013 elections as cases. Both countries experienced serious economic hardships after the global credit crunch in 2008, with International Monetary Fund bail-out for Iceland and in the case of Greece, a bail-out by the so called Troika, i.e. the European Union, the European Central Bank and the International Monetary Fund. Both had diplomatic disputes with the European Union or single countries within the European Union because of the crisis, as well as political turmoil with voters shifting their alliances between parties and the emergence of new parties. In both countries there are political parties that have run on a platform of populist promises about how to deal with the domestic economy and foreign creditors and they have earned an unexpected large share of the votes. In both countries there were very popular voting advice applications that were used by a significant part of the voting populations (7.4% in Greece, 15.5% in Iceland). In our paper we use data from these voting advice applications to show that there is a significant group of voters in both countries who appear to have made their vote choice not on the basis of an overall ideological congruence, but because these parties have been promising a way to deal with the financial crisis that would remove most of the weight from the shoulders of the voters.

### **The Politics of U.S. Asians**

Lydia Saad, *Gallup, Inc.*

Andrew Dugan, *Gallup, Inc.*

The high proportion of Asians voting Democratic in recent presidential elections -- 73% voted for Obama in 2012 -- has prompted a great deal of attention, especially as Asian-Americans have earned the "fastest growth" distinction of any U.S. ethnic or racial group. There is also considerable speculation as to why Asian-Americans, who typically enjoy much higher income levels than blacks or Hispanics, nonetheless align with these groups, politically. Among the many explanations proffered are the Democratic Party's concerted outreach to Asian voters, the Democratic Party's more immigrant- and minority-friendly political stances, and the Democratic Party's more liberal social issue platforms relative to the Republicans'. However, another factor may trump all of these in importance: religion. Unlike with blacks and Hispanics, whose party allegiance slants heavily Democratic regardless of religious identity, Asians' party leanings demonstrate high variability according to religious preference. This ranges from high proportions of non-Christian and atheist Asians identifying as Democratic, to a more even split in party



identification among Protestant Asians. Additionally, whereas most blacks and Hispanics identify as Christian, fewer than half of Asians do. Thus, the influence of atheists and non-Christians on overall political preferences is much greater for Asians than for blacks or Hispanics. This paper will use Gallup Daily tracking data from 2008 through 2013 to closely analyze the politics of Asians relative to blacks, Hispanics, and non-Hispanic whites. The paper will review the importance of religion as a determining factor in Asians' strong Democratic orientation both alone, and in combination with other demographic characteristics, and attitudes such as political ideology. Gallup tracking involves interviews with 350,000 national adults each year, including approximately 7,000 Asians, representing an unparalleled dataset of nationally representative Asians from which to glean new insights about the politics of this increasingly important voting bloc.

### **Attitudes and Ambivalence toward Political Candidates: An Asymmetric Nonlinear Approach**

Drew Allen, *City University of New York*

Individuals' attitudes toward candidates directly influence voting behavior and therefore have important implications for candidates, campaign strategies, and political institutions themselves. Citizens often harbor diverse sets of feelings and beliefs about presidential candidates that influence their overall attitude toward them. Zaller (1992) stresses that individuals do not typically possess "true attitudes" toward subjects, but rather combine information and feelings about candidates to form and express a net attitude. The goal of this paper is to evaluate how individuals combine varying and sometimes conflicting beliefs toward candidates into overall attitudes toward them. Using an asymmetric nonlinear model approach first described by Holbrook, et al. (2001), this study evaluates individuals' feelings towards presidential candidates from 2000 through 2012 and attempts to answer the following question: How are individuals' attitudes toward candidates in these presidential elections best predicted? The paper uses pooled National Election Studies data to examine the asymmetric nonlinear approach among various elections, parties, and subgroups of individuals to determine this approach's generalizability. Overall, the asymmetric nonlinear model describes attitudes toward presidential candidates significantly better than a traditional symmetric linear approach. These findings have important implications for the study of ambivalence. Many researchers suggest that individuals who possess similar numbers of positive and negative feelings about a candidate are ambivalent toward them. However, an asymmetric nonlinear approach to ambivalence reveals that this may not always be the case.

## **Methodological Briefs: Questionnaire Design**

### **Measurement Properties of Brief "Big Five" Personality Instrumentation**

Matthew DeBell, *Stanford University*

Ted Brader, *University of Michigan*

Simon Jackman, *Stanford University*

Catherine Wilson, *23andMe*

The "Big Five" personality traits are the subject of a huge literature in psychology and are increasingly of interest in allied disciplines, having recently been measured in the American National Election Studies and proposed for the Comparative Study of Electoral Systems. Much of the psychology literature employs extensive multi-item scales that are too long for representative sample surveys. The Ten Item Personality Inventory (TIPI) has made Big 5

measurement practical in more settings, including representative surveys. However, TIPI's design has suboptimal features: agree-disagree response options, numeric instead of verbal response options, and multiple items on the same page. In this paper we report the results of an attempt to improve personality measurement by editing the questions to fix these problems. We compare the original version of TIPI to a revised version and comprehensively assess the quality of the resulting data (from surveys conducted by the American National Election Studies in 2012 and 2013) on several dimensions: completion time, item nonresponse, paired item reliability, and construct validity. This paper presents the first tests of construct validity for all five personality items in this edited form of TIPI. We find that completion time is the same and item nonresponse rates are both low, while reliability and construct validity for the revised TIPI are better than the original version by most measures. The results show how better personality data can be obtained at no additional cost by optimizing questionnaire design.

## **Using Paradata to Predict Interviewers' Likelihood to Take Question-Reading Shortcuts**

Jennifer Kelley, *University of Michigan*

Ziena Mneimneh, *University of Michigan*

A well-known tenet of survey question design is to keep it 'short and simple'. The main objective of this tenet is to improve question comprehension and reduce respondent cognitive burden. However, there is little research on the length and complexity of survey questions and interviewer reading behavior. It is possible that longer and/or more complex questions could lead interviewers to take shortcuts such as paraphrasing, only asking key phrases or words, or skipping the question entirely. When does the question become too long and/or complex that it triggers such shortcuts? Are there certain types of questions (e.g. attitude or factual; close-ended or open-ended) that are more susceptible to shortcuts than others?

This presentation will attempt to answer the above questions by using paradata from two face-to-face surveys. A sample of questions of varying types and length will be selected for analysis. For each question, a measure of its length and complexity will be established. Question length will be determined by counting the number of words. Question complexity will be measured by evaluating the number of concepts the respondent is being asked to hold in their mind to formulate their answer. To determine whether a short-cut was taken, the threshold time will be compared to the administration time (measured by timestamps). The threshold-time is defined as the fastest time a question can be read and still preserve respondent comprehension. For each question, a binary outcome will be created based on whether the threshold-time is less than the administration time or not. This outcome will then be regressed on question length, question complexity, whether the response option is part of the question itself and the type of question. The presentation will conclude by providing recommendations for questionnaire designers on the optimal question length beyond which interviewers might be engaging in question-reading shortcuts.

## **The Utility of Unbalanced Bipolar or Shortened-Range Response Scales in Testing Change in Attitudes, Opinions and Ratings: a Simulation Study**

Sonja Ziniel, *Boston Children's Hospital, Harvard Medical School*

Al Ozonoff, *Boston Children's Hospital, Harvard Medical School*

Recommendations for textual content of good attitudinal, opinion or rating questions include the use of bipolar ordinal response scales with balanced number of response options in either direction such as a 5-point scale ranging from "Strongly disagree" to "Strongly agree" and

centered on “Neutral”, assuming that balanced response scales avoid directional bias, irrespective of including a middle/neutral option. This approach has been shown effective for attitudinal questions where an assumption can be made that attitudes in the survey population are distributed across all response categories approximating an underlying latent normal distribution.

We extend these investigations to situations where survey instruments are developed to measure change over time of attitudes, opinions, or ratings that are highly skewed. Questions with highly skewed response options are especially undesirable if they are designed to measure change over time. In these settings the response distribution will exhibit bottom/ceiling effects which tend to reduce the statistical power to detect true changes in the population over time. We propose using unbalanced bipolar response scales or unipolar scales that do not span the entire range of attitudes, opinions and ratings in these situations. Furthermore, we compare the performance of these scales against the “standard” four-point and five-point responses in the context of pre-post hypothesis tests of mean changes in attitude. Our evaluation uses simulated latent attitude responses under a pre-post mean shift. We examine the usefulness of different unbalanced response scales when the underlying latent response distribution is asymmetric and non-normal. We aim to provide theoretical foundations for the development of appropriate unbalanced bipolar or shortened-range response scales when measuring already skewed attitudes, opinions or ratings that can be expected to intensify.

## **A General Survey Measure of the Need for Closure**

Eike Mark Rinke, *University of Mannheim*

A 5-item short form of the Need for Closure Scale (NFCS-5) for use in general population surveys is proposed. Recently, interest in the motivational and cognitive underpinnings of political information processing, attitudes, and behavior has resurged (e.g., Jost, Glaser, Kruglanski, & Sulloway, 2003), especially interest in the political consequences of dispositional need for cognitive closure (e.g., Chirumbolo, Areni, & Sensales, 2004). The Need for Closure scale (NFCS) measures an individual’s motivation to “seize and freeze” on beliefs that offer simplicity, certainty, and clarity and has been well-validated (Webster & Kruglanski, 1994). However, the original 41-item scale is obviously too lengthy for use in general population surveys, and the same is true for its validated shorter versions. A previous attempt to validate an adapted 5-item version of the NFCS in an ANES pilot study was not successful (Federico, Jost, Pierro, & Kruglanski, 2007).

The NFCS-5 proposed here is adapted from prior validation analyses by Roets & Van Hiel (2011) and was validated in three studies. Study 1, based on an online survey of a US-American student sample ( $N = 382$ ), and Study 2, based on an online survey of a US-American general population sample ( $N = 1,023$ ), demonstrated adequate internal consistency ( $\alpha \geq .72$ ) and the expected factorial structure of the NFCS-5; its consistency with a validated 15-item NFCS; and its predicted association with a relevant personality measure (need for cognition). Study 3, based on a online-surveyed general population sample from the German Longitudinal Election Studies (GLES;  $N = 1,049$ ), replicated reliability ( $\alpha = .71$ ) and factorial structure of the NFCS-5; and demonstrated its criterion validity in predicting measures of political conservatism and group-centrism. The NFCS-5 was economic with administration times ranging from 37 to 44 seconds across Studies 1-3.

## **`Good Respondent, Bad Respondent'? Assessing Response Quality in Internet Surveys**

Antje Kirchner, *Survey Research and Methodology (SRAM)*

Rebecca J. Powell, *Survey Research and Methodology (SRAM)*

Much attention has been devoted to the assessment of survey response quality in the absence of validation data, more recently this includes a focus on paradata (Kreuter 2013). While existing empirical studies typically analyze only individual indicators as measures of response quality, more research is needed to understand the latent construct of 'response quality' by combining multiple indicators. This study examines response quality--using multiple indicators--through cross-classified structural equation modeling.

We start by providing an extensive literature review of previously used indicators of response quality in web surveys. The main research question is whether we can use multiple response quality indicators--such as nondifferentiation, response latencies, or the number of answer changes--to capture the latent construct response quality in web surveys. This will allow us to distinguish so-called 'good' from 'bad' survey respondents. We also assess whether this composition differs by device type (e.g. computers, tablets, cell phones) that was used by a respondent to complete the survey.

The analyses are based on the internet component of the Gallup Panel using cross-classified models. Initial results suggest that there is one common factor 'response quality' relating response latency, answer changes, page break-offs, nondifferentiation, 'don't know' responses, and item nonresponse. Although all factor loadings are significant, nondifferentiation and answer changes particularly contribute to the latent factor: the less nondifferentiation and the fewer answer changes, the higher the factor score indicating a 'better' respondent. Surprisingly, page break-offs and the percentage of 'don't know' responses contribute less to the measurement of response quality. This paper further discusses the implications of using one latent factor compared to single indicators to measure response quality.

## **The Effects of Cover Images on Participation and Reports in Mail Surveys**

Mathew Stange, *University of Nebraska-Lincoln*

Cover images brand surveys in ways that may influence participation and possibly reports to questions, which is why researchers often choose cover images for mail surveys that relate to a survey's topic and goals (Dillman, et al. 2009). One goal of this study was to identify lesbian, gay, and bisexual (LGB) individuals and couples within a general population survey. This required methods to encourage their participation because LGB people and couples sometimes are reluctant to identify themselves in surveys because of social stigma (Gates 2010). One possible way to encourage participation is through a cover image that brands the survey as inclusive of LGB individuals by drawing upon a sense of group identity (Borgerson, et al. 2006). However, the inclusive cover image can potentially have negative effects, such as reducing non-LGB individuals' participation and influencing reports to questions. In this paper, I examine results from an experiment of three cover image versions: 1) a mix of images of traditional families and individuals displaying themselves in typical gender ways, 2) a mix of images of LGB individuals and families and heterosexual individuals and traditional families, and 3) no cover image. The experiment was embedded in the 2013 Nebraska Annual Social Indicators Survey, a general population mail survey of Nebraskans (n=1,608). I examine differences in response rates and make-up of the completed respondent pools across each condition, including the number of LGB individuals and couples. I also examine if the content of the cover

images influenced reports to political questions among all respondents. Preliminary analyses show that the participation rate was highest in the no cover image condition (29%), followed by the LGB-inclusive condition (27%), with the lowest participation rate among the cover image condition of traditional families (25%). My paper will end with recommendations on using cover images in mail surveys.

## **The Effects of Response Option Changes on Response and Data Quality**

Cong Ye, *American Institutes for Research*

Jill Watson, *American Institutes for Research*

Rebecca Medway, *American Institutes for Research*

Respondents may draw inferences from the structure of response options. As Schwarz, Hippler, Deutsch, and Strack (1985) argue, respondents consider the middle option as the typical value when answering a frequency question; for example, higher reports of television viewing behavior resulted from a scale with high frequency options (“up to 2.5 hours” to “more than 4.5 hours”) than from a scale with low frequency alternatives (“up to 0.5 hour” to “more than 2.5 hours”). Similar results have been found with reporting of other behaviors. In addition, respondents may choose a response that they are not certain about in a forced-choice question, but would instead select a “no opinion” response if this option were explicit (e.g., Bishop et al. 1986). This raises the question of whether a noncommittal option might change the distribution of substantive positions.

AARP’s Experience Corps (EC) program places adults aged 50 and older into public schools to tutor and mentor students in reading and literacy. Every year, the program includes approximately 2,000 EC Members who work with about 20,000 students in about 20 cities. Students’ teachers and EC Members are asked to complete self-administered questionnaires to report students’ performance and other aspects of the program. The survey questionnaires had significant changes between the 2011-12 and the 2012-13 versions which allow us to investigate questions about the impact of the response option structure. Do unbalanced options (such as a skewed midpoint) lead to more extreme answers and change the percent reporting above the midpoint? Is a yes-no option format more likely to lead to acquiescence than a yes-no-maybe option format? How do the two changes affect item nonresponse? Using the original 2011-12 questions to establish a baseline for the evaluation, we will investigate differences in responses to the modified survey items to address these questions.

## **Using Motivating Prompts to Increase Responses to Open-ended Questions in Mixed-mode Surveys: Further Evidence on Where the Prompt Should Be Placed**

Glenn D. Israel, *University of Florida*

Getting respondents to provide high quality information to open-ended questions in self-administered surveys is challenging. Evidence shows visual and verbal design elements play a role in response behavior. Regarding visual design, providing an “optimal” size answer space contributes to higher item response and longer answers in mail and web surveys (Israel, 2010; Smyth et al., 2009). Likewise, including motivating statements in the stem has improved response quality in Web surveys (Smyth et al., 2009). Finally, mode impacts responses, with Web surveys eliciting longer answers than mail surveys. Given interest in mixed-mode surveys, I explore the effect of adding a motivating prompt to open-ended questions to assess impacts on item response rate and response length for mail and Web modes. Further, I test whether placing the prompt at the beginning or end of the question affects responses. Data from a survey of Cooperative Extension Service clients in 2012 and 2013 are used for the study. The

importance prompt consistently increased the item response rate for the question about improving Extension's services but it had mixed results on the description question asking clients about getting information, its use and the result. I also found that the importance prompt increased the number of words in answers provided by respondents for the improvement question over having no prompt. This effect occurred for the prompt placed either at the beginning of the question or at the end. The importance prompt did not affect response length for the description question. Web responses were longer than mail, independent of the prompt for both questions. The findings suggest there is some benefit to occasionally using a motivating prompt but it is unclear when it will be helpful.

## **U.S. Public Opinion and the Economy**

### **Community Development Stakeholder Data: Exploring Local Community Conditions, Economic Inclusion and Emerging Issues**

Barbara J. Robles, *Board of Governors of the Federal Reserve System*

In an effort to mitigate knowledge gaps related to economic growth and inclusion at the community level, the Federal Reserve System began an exploratory, pilot process of collecting quantitative and qualitative data from community stakeholders serving low-to-moderate income (LMI) consumers, families and communities using web-based surveys. Community stakeholders such as: community development financial institutions, credit unions, community banks, NGOs, public sector agencies, faith-based organizations, universities and community colleges are often the central entities of community economic development in LMI locales. Recognizing that anecdotal information can provide a narrative contextualizing local change and shed light on numeric survey responses, data was collected at the community level by regional districts. In addition, community level data was collected by administering a national web-based survey in order to triangulate findings and provide comparison benchmarks. We identified local and economic inclusion metrics that captured trends and geographical dispersion. In monitoring the recovery capacity of low-to-moderate income communities, we gathered information on a variety of economic inclusion categories: financial well-being, access to credit, availability of jobs/employment, affordable housing and other measures at the local level. By employing a mixed methods approach, we have been able to analyze the information from a variety of quantitative and qualitative data visualization approaches. Such analysis provides an opening for a novel integration of micro-data to assess regional and macro impacts.

### **The New American Economy**

Erica Seifert, *Greenberg, Quinlan and Rosner Research*

Over the last 5 years, Democracy Corps has conducted scores of surveys, in-depth-interviews, and focus groups on Americans' personal economic experiences. Their personal economies have shifted significantly over the course of the last 5 years-- from devising makeshift strategies to make ends meet during tough times-- to now permanently downsizing their expectations and definitions of the American Dream. Elite economic indicators-- GDP, Labor Department jobs numbers, and closing bell figures-- have almost nothing to do with the pocketbook-level grocery store indicators Americans live and measure every day. When they hear about new jobs and jobs growth they do not cheer, but instead ask, "How do these jobs stack up to the cost of living?" This downsizing now seems permanent and pervasive. Our questions dig deeply into these micro-economies to measure and examine the real economy-- one where Americans

have very different assumptions about life chances, opportunity, income, and equity, than elites might expect.

### **The Course and Correlates of Consumer Sentiment 1985-2014**

Gary Langer, *Langer Research*

Langer will discuss the course of consumer sentiment during the last several years as measured in the 28-year-old weekly Bloomberg Consumer Comfort Index, including attitudes across groups, an evaluation of some of the key correlates of consumer sentiment and a discussion of its role in predictive models of other economic indicators. He'll cover the political as well as economic implications of consumer sentiment, with an eye toward the 2014 midterm elections, and suggest directions for future research.

### **The Current and Future Impact of Millennials on the U.S. Economy**

Paul Taylor, *Pew Center for the People and the Press*

This presentation on Millennials will examine a mix of economic and attitudinal data. The economic data will compare Millennials with older adults on a range of key measures – income, wealth, poverty, debt, unemployment and household formation. Where possible, it will compare the economic situation of Millennials today with that of older adults back when they were the age that Millennials are now.

The attitudinal data will draw on Pew Research surveys that explore Millennials' attitudes about their jobs, their careers, the cost and value of their education, their expectations for their future economic well-being, and their attitudes about retirement; again, we'll make comparisons with older age groups.

### **Adjusting, Organizing, and Fortifying Personal Finances in a New American Economy: 2009 to 2014**

Janet L. Streicher, *Citibank*

Citi has been conducting consumer polls since the beginning of the crisis in US financial and credit markets that emerged in the fall of 2008. This paper will focus on consumers' adaption to the economy, their financial challenges and resourcefulness as they manage through a sluggish national economy that continues to disappoint.

The National Citi Economic Pulse survey has been conducted 11 times to date, but the presentation will include 12 by the time of the AAPOR conference in May, 2014. Each survey is composed of an index and then we select topics germane to consumer sentiment on the economy and their personal finances. The Pulse itself combines eight survey questions into a single measure of overall economic status and future outlook and will be briefly reviewed for variation among important groups such as income, generation, class, and other factors. The key focus, however, of this presentation will be on the how Americans are moving their financial life forward in this new economy. Key topics will explore changes in spending, borrowing, saving and investing behavior, the impact of mobile banking on personal financial management and control, and resiliency in stretching income and available funds to achieve financial stability.

## Developments in Cognitive Interviewing Methods

### Crowdsourcing in the Cognitive Interviewing Process

Joe Murphy, *RTI International*

Michael Keating, *RTI International*

Jennifer Edgar, *U.S. Bureau of Labor Statistics*

Crowdsourcing, or tapping into the collective intelligence of the public to complete a task.” (King 2013), holds promise as an alternative or supplement to traditional cognitive interviewing methods. By providing researchers access to a much wider range of potential participants than is typically available in lab-based methods, crowdsourcing has the potential to create a sample more demographically or geographically similar to the survey population of interest. Research (e.g. Keating et. al 2013) has found that crowdsourcing platforms can provide a viable alternative to traditional data collection. In addition, exploratory work looking at the use of self-administered cognitive interviewing methods (e.g. Edgar, 2013) has found promising results in terms of the quality and quantity of information collected.

Using crowdsourcing to recruit participants to complete cognitive interviewing-type studies online, may address many of the existing challenges facing researchers today. Participants can be recruited much faster, with far lower researcher time required. Participant incentives and the overall cost of the study (especially when including recruiter and interviewer time) can be lower for crowdsourcing methods. These time efficiencies allow for larger sample sizes than are typically possible, or could allow for more iterative testing on question drafts than is typically done.

This study compares three crowdsourcing methods (TryMyUI, Facebook and Amazon Mechanical Turk) to traditional lab-based cognitive interview. Participants in all three groups, as well as those completing traditional lab-based cognitive interviewing, completed several tasks aimed at collecting information about their comprehension and response process for survey questions. We look at recruiting results, as well as the quality and quantity of the substantive data collected. Advantages and disadvantages of each crowdsourcing method are compared, as well as the overall implications for this approach.

### Reliability in Qualitative Research: Now What Does that Question Mean to You?

Darby Steiger, *Westat*

David Cantor, *Westat*

Aaron Maitland, *Westat*

Martha Stapleton, *Westat*

Measures of reliability are often conceived of as a measure of quality in quantitative research, in which standardized questionnaires can be re-administered to the same or independent populations to measure consistency in responses. It is more unusual, however, to measure reliability of non-standardized qualitative data, and this is often a criticism of qualitative research (Willis, 2005). Reliability is often discussed in terms of inter-rater reliability of quantitative data, rather than the reliability of the qualitative data (Daly et al, 1992; Armstrong et al, 1997). For cognitive interviews, an important question is whether the qualitative conclusions about design flaws are reliable. This paper reports on the findings from a series of cognitive interviews in 2013, in which 23 women answered questions about their unwanted sexual experiences over the past 12 months, and participated in the same cognitive interview two weeks later. Respondents reported a total of 88 incidents of unwanted sexual experiences in the first



interview and only 73 incidents in the second interview. The paper will describe key reasons for the changes in answers that map onto common theories of cognitive processing (Tourangeau et al, 2000). When asked to explain their discrepant answers in the second interview, respondents reported different interpretations of the questions in the second interview (comprehension), recalling new details of incidents between the first and second interview (retrieval), thinking about their experiences differently (judgment), and deciding on a different way to answer the question (mapping). These results suggest that many of the changes are due to respondent conditioning, which generally improved the quality of the data at the second interview. The re-interview design, therefore, seemed to provide measures of both validity and reliability. Implications of using re-interview designs when pre-testing questions are discussed.

### **Enumerating Persons With No Address in an Address-Based Census**

Katherine R. Kenward, *Research Support Services, Inc.*

Eleanor R. Gerber, *Research Support Services, Inc.*

Alisu Schoua-Glusberg, *Research Support Services, Inc.*

Patricia Goerman, *U.S. Census Bureau*

Murrey Olmstead, *RTI International*

In an effort to enumerate every person in the United States, the Census Bureau has created several instruments which are designed to count people who may not be reached by ordinary procedures. One such instrument, the Be Counted form, is distributed in public locations like libraries and convenience stores. Recent cognitive and usability research was undertaken to examine the effectiveness of the Be Counted instrument in both paper and digital modes. The object of this paper is to examine how respondents who are not attached to a single address interact with the experimental Be Counted instrument. This includes two subgroups through which the same individual may cycle at various times: persons tenuously attached to households and persons who stay only in street locations or emergency shelters. This paper focuses on a question designed to screen for homelessness, as well as the people's responses to subsequent survey items that require an address or imply that the respondent has an address. Issues addressed include: 1) the effect of the self-perception of "homelessness" on the provision of adequate responses; 2) How well the information provided by respondents on the forms reflects their verbal descriptions of their living situation and 3) Homeless respondents' interpretation of questions which assume that they have a fixed single address. Possible alternative ways of collecting geographical information about this population are considered as well.

### **Qualitative Testing of the 2016 Canadian Census of Agriculture Questionnaire**

Diane Fournier, *Statistics Canada*

A significant shift in data collection methodology has taken place in recent years at Statistics Canada as the Agency transitions to using web questionnaires as the primary collection tool. The challenges that have ensued from using web questionnaires have modified testing methodologies to include a mixed-mode approach.

For the first time at Statistics Canada, testing in an environment mimicking Census Day for Canadian farmers took place during the fall in 2013. Multiple phases of testing took place for the 2016 Census of Agriculture starting with the testing of new concepts and questions through the use of paper questionnaires and prototypes. As testing progressed and questions solidified, the testing strategy changed to testing a web-based version of the questionnaire. The last two phases of testing took place on-site with farmers across Canada using their own computers and

Internet access. Questions that tested well in paper form needed to also test well for participants online. Analysing the quality of questions, instrument instructions and survey navigation were all part of the objectives of this study but more than that, the whole user experience was being observed and evaluated.

Changing the testing strategy from paper to online also changed the testing experience for both the interviewers and participants. The advantages and disadvantages of the testing environment for the Census of Agriculture will be examined and discussed.

## **Using Cognitive Interviews to Test Behaviorally Specific Questions on Rape and Sexual Assault**

David Cantor, *Westat*

Darby Steiger, *Westat*

Shannan Catalano, *Bureau of Justice Statistics*

Kay Ricci, *University of Nebraska – Gallup, Inc.*

Reanne Townsend, *University of Maryland - JPSM*

National estimates of rape and sexual assault (RSA) differ by factors of between 5 to 10 (Bialak, 2013). One of the survey design features associated with this variation is the use of “behaviorally specific” questions that use explicit language to detect and classify self-reports as RSA. Surveys that use this strategy tend to produce the highest rates of RSA. This paper addresses three research questions related to this methodology: 1) how do respondents interpret key terms used when classifying events as RSA (e.g., ‘force’, ‘unable to consent’, ‘against your will’) 2) what types of incidents are reported from these questions and 3) how does respondent interpretation affect the classification of the incident as RSA?

The analysis is based on 35 cognitive interviews conducted in 2013 with women who were pre-screened as having had an unwanted sexual experience. The cognitive interviews tested a two-stage questionnaire designed to identify and classify incidents of RSA. The first stage administered screening questions drawn from existing surveys that use a behaviorally-specific approach to measurement. The second stage administered follow-up questions about the details of the incident, as well as collecting a verbal description of the incident by the respondent. Based on the results from the first stage, incidents reported by 26 of the women were classified as a rape. However, based on additional information from the verbal descriptions and second stage, only half of these 26 incidents were classified as a rape. Key terms such as ‘force’, ‘against your will’ and ‘unable to consent’ were not uniformly interpreted as intended by the survey designers. The role of alcohol, and its relationship to legal definitions of rape, also led to ambiguous classifications. These results highlight the complexity of measuring RSA and suggest that multiple questionnaire items are needed to accurately describe events as complex as RSA.

## **The Dawn of ACA Implementation: Research Challenges and Early Data**

### **Measuring Health Coverage in Surveys Post-Reform**

Joanne Pascale, *U.S. Census Bureau*

For more than a decade, the U.S. Census Bureau has been conducting research on reducing measurement error in its surveys that measure health insurance, culminating in a redesign of

the Current Population Survey (CPS) that was tested in a large-scale content test in March of 2013. Results indicated that the redesign produced more accurate and more useful data than the conventional questionnaire, and the Census Bureau expects to adopt the redesign as a production instrument in the March 2014 CPS and beyond. In the meantime, the Affordable Care Act (ACA) was passed in 2010, and in 2011 the Census Bureau began investigating how to modify its surveys to capture coverage post-reform. One of the most significant departures from conventional sources of coverage under the ACA is the “Health Insurance Exchange” (now more commonly called the “marketplace”) – a state-level marketplace of health insurance options for individuals and small businesses. In 2006, Massachusetts passed state-level health reform, including the marketplace, on which the ACA was largely modeled. This provided a unique opportunity to conduct qualitative research with state residents known to be enrolled in the marketplace to explore the many pathways of enrollment, the language and terminology residents used when describing their coverage, and ultimately to develop standardized questions for capturing marketplace participation and subsidization. These questions were developed and tested within the context of both the Current Population Survey (CPS) redesign and the American Community Survey, thereby providing some baseline findings for other federal and state surveys. This paper will present highlights from the 2013 test of the CPS redesign and from the Massachusetts research on measuring coverage through the marketplace. It will also address plans for ongoing research to assess both the CPS redesign and the questions on health reform once reform goes into effect in 2014.

### **Familiarity with ACA Provisions and Understanding of Personal Impact on the Eve of Health Reform**

Alisha Baines Simon, *Minnesota Department of Health*

Stefan Gildemeister, *Minnesota Department of Health*

Kathleen Thiede Call, *University of Minnesota*

States have taken disparate approaches to the Affordable Care Act (ACA) and its implementation. For this reason, while national polls can provide a general picture, limited data on state-specific reaction to the ACA makes it difficult to predict how the ACA may work across the country. Minnesota has generally embraced the ACA and its reforms as evidenced by an early expansion of Medicaid in 2011 and full expansion in 2014, as well as a state-based exchange, called MNsure. To better understand familiarity with the ACA, new questions were added to the biennial Minnesota Health Access Survey (MNHA) in 2013. The MNHA is a dual-frame RDD/cellular telephone survey of more than 11,000 Minnesota residents, which provides estimates of health insurance coverage and access in the state. One question (from the Kaiser Family Foundation Health Tracking Poll) asked if the respondent felt that they understood enough about the health reform law to understand how it would affect them. The survey also included questions asking about familiarity with five provisions of the law: Medicaid expansion, insurance exchanges, premium subsidies, guaranteed issue and the individual mandate. This paper explores how groups most likely to be impacted by the law (the uninsured, and those eligible for Medicaid and insurance subsidies on the exchange) understand the law, and if they are familiar with parts of the law most likely to be applicable to them. Familiarity with specific provisions is not uniform, with higher familiarity with the individual mandate and guaranteed issue, and lower familiarity with MNsure and premium subsidies. Further, familiarity differs by insurance status, with the uninsured having the lowest familiarity. Nonetheless, understanding of the ACA's impact, as well as familiarity with its parts increased over the survey field period (from August to November, 2013), which straddled the launch of MNsure open enrollment.

## **Tracking the Experience of U.S. Adults in the Affordable Care Act's Marketplaces: The Commonwealth Fund Affordable Care Act Tracking Surveys, 2013-2014**

Sara R. Collins, *The Commonwealth Fund*

Petra W. Rasmussen, *The Commonwealth Fund*

Robyn Rapoport, *SSRS/Social Science Research Solutions*

Open enrollment through the Affordable Care Act's insurance marketplaces extends from October 1, 2013 through March 31, 2014. In order to provide data regarding early experiences with the new insurance options, the Commonwealth Fund (Fund) and Social Science Research Solutions (SSRS) are conducting a series of four surveys to track people's awareness of, and experiences with, the new health insurance marketplaces just prior to and during the open enrollment period. Fielded from early July through September 30th, the first survey of 6,132 adults ages 19 to 64 found that only 2 of 5 adults were aware of the marketplaces or of premium subsidies, with awareness lowest among those who are uninsured or earned low incomes. In October, the Fund and SSRS interviewed a sample of 682 adults who were uninsured or had a health plan through the individual market. The survey found that awareness climbed substantially from the first survey and that 17 percent of those potentially eligible had visited the marketplaces. The data reflect the rough rollout of the federal marketplace in October with large shares of adults who visited the marketplaces indicating difficulty comparing plans and determining whether they could afford the plans. Similar surveys of the potentially-eligible population will be conducted in December 2013 and April 2014. The paper will (1) examine awareness and use of the ACA marketplaces over time, (2) discuss challenges associated with creating survey questions to capture these experiences, and (3) offer interpretations of the data trends. It will also examine the use of survey findings by the media and policy makers.

## **Interviewing California's Uninsured on the Eve of Open Enrollment: The Kaiser Family Foundation 2013 Baseline Survey**

Mollyann Brodie, *Kaiser Family Foundation*

Liz Hamel, *Kaiser Family Foundation*

Claudia Deane, *Kaiser Family Foundation*

In Washington, the 2010 Affordable Care Act (ACA) is portrayed primarily as a political weapon being wielded by both major political parties in their struggle to control the branches of government. But in California, the nation's largest state with the concomitantly largest number of uninsured Americans, the ACA is an on-the-ground policy change of the greatest magnitude. In the summer of 2013, as the state government continued its active effort to prepare for the October 1 open enrollment period of its new insurance exchange (Covered California), the Kaiser Family Foundation fielded a representative, probability based sample survey of 2,000 uninsured Californians designed to capture the experiences of this population during an unprecedented opportunity to gain coverage. Because eligibility for the law's extended coverage benefits –via either the Medi-Cal program or the insurance exchange -- is tied both to income (as measured by Federal Poverty Level benchmarks) and to citizenship/legal resident status, the survey faced the twin challenges of measuring these two sensitive issues. This paper will present an assessment of the challenges faced in ascertaining eligibility for the new benefits via survey and the success of that effort. It will also report on the substantive findings of this unique effort, including levels of awareness of the upcoming policy changes among California's eligible uninsured population, presumptions of eligibility, and interest in becoming insured. The paper will also consider the expectations of the undocumented uninsured population, a group that may expect help with coverage but will not be eligible under the law. Finally, as this is the first of four waves of the California Uninsured survey, we will also report on progress on the project's

second wave, to be fielded in the spring, and the difficult work of garnering accurate measures of the enrollment experience.

**Friday, May 16**  
**3:15 p.m. – 4:15 p.m.**  
**Poster Session 2**

**1. The Structure of Foreign Policy Attitudes Redux: Cross-National Evidence**

Timothy B. Gravelle, *University of Essex & PriceMetrix Inc.*

A recurring debate in the field of public opinion toward foreign policy relates to the number, content, and cross-national equivalence of the overarching concepts that serve to organize mass public attitudes toward foreign affairs. Put more succinctly, how many factors are there, what should we call them, and are they the same or different in different countries? This paper brings new data from a series of Economic and Social Research Council-sponsored surveys to bear on these questions. These surveys, conducted in parallel in a number of advanced industrialized countries, provide an expanded set of foreign policy-related survey items that are analyzed using the tools of exploratory and confirmatory factor analysis. The paper argues for a more multi-faceted (and multi-factor) view of how mass publics structure their attitudes toward foreign affairs.

**2. Use of the Massachusetts Health Insurance Exchange in Three Distinct Boston Neighborhoods**

Jessica L. LeBlanc, *Center for Survey Research, University of Massachusetts Boston*  
Anthony M. Roman, *Center for Survey Research, University of Massachusetts Boston*  
Philip S. Brenner, *Department of Sociology, University of Massachusetts Boston*  
Naa Oyo A. Kwate, *Rutgers University*

With the advent of the Affordable Care Act, many states are installing public health insurance exchanges to offer health insurance products to the public for the first time. With this in mind, a look into who uses these exchanges and how they are used in a state with a well-established exchange may be of interest. In 2006, years before the Affordable Care Act, The Commonwealth of Massachusetts created the Massachusetts Health Connector (the Connector). In 2013, a study was conducted to investigate, among other topics, Boston residents' experiences getting health insurance through the Connector. An address-based sample of residents from three Boston neighborhoods, distinct in racial and economic makeup was surveyed by mail using a three-contact protocol, with an overall response rate of 38% (AAPOR RR4). While some study results were quite predictable, others differed from initial expectations. Neighborhoods with higher median incomes had fewer residents use the exchange to purchase health insurance, and of those that did use the exchange, fewer qualified for a plan with no premiums. For those who do pay premiums, the percentage of people receiving some form of subsidy goes down as the median income of a neighborhood goes up. These results are not surprising. However, even in the highest median income neighborhood, over 30% of those getting insurance through the exchange qualify for some form of subsidy. Also, despite that none these neighborhoods are considered "low-income," in two of the three neighborhoods, about a third of those using the exchange qualify for getting insurance with no premium at all. Finally, in all three neighborhoods, the majority of residents still get health insurance through alternative means, with the highest rate of Connector use in any of the neighborhoods at 17%. Breakdowns of use of the Connector by race, age and gender are also examined.

### **3. Gender and Abortion Attitudes: Making Sense of the Null Relationship**

Steven E. Barkan, *University of Maine*

Despite reason to believe from an interest group perspective that women should be more likely than men to support legal abortion, much research finds instead that women and men hold similar views on this issue. This null relationship continues to puzzle scholars of abortion attitudes. In an attempt to unravel the puzzle, this paper argues that gender differences in religiosity help explain the lack of a gender difference in abortion attitudes, with religiosity acting as a suppressor variable for the theoretically expected relationship between gender and support for legal abortion. The paper tests this hypothesis with data from the 2012 General Social Survey. Supporting the hypothesis, the expected gender difference in support for legal abortion emerges when religiosity is controlled in multivariate analysis. This result indicates that religiosity is indeed suppressing women's greater support for legal abortion that would be anticipated from an interest group perspective. Final remarks outline directions for future research suggested by the analysis.

### **4. Cognitive Interviews of Veterans for the Gulf War Era Cohort and Biorepository Project**

Julie C. Linville, *Social & Scientific Systems, Inc.*

Edward E. Gaunt, *Social & Scientific Systems, Inc.*

Karen Goldstein, *Department of Veterans Affairs Medical Center*

Rebecca B. McNeil, *Department of Veterans Affairs Medical Center*

Dawn T. Provenziale, *Department of Veterans Affairs Medical Center*

Catherine M. Thomas, *Department of Veterans Affairs Medical Center*

Corrine I. Voils, *Department of Veterans Affairs Medical Center*

Specific methods used in and results obtained from cognitive interviews are rarely included in the literature. Many survey methods articles indicate that cognitive interviews were conducted, but do not provide in-depth detail on respondent selection, the procedures used, the sequence of procedures, how the observations were recorded, how the observations were interpreted or the resulting survey instrument changes. In an effort to encourage more complete reporting of cognitive interview methods and results, we will discuss cognitive interviews conducted as part of the Gulf War Era Cohort and Biorepository, a project sponsored by the Office of Research and Development in the Department of Veterans Affairs (VA). The project will enroll approximately 30,000 Veterans who served in the uniformed services between 1990 and 1991 and consists of a mailed self-administered questionnaire completed by the Veteran in their home, followed by a visit from a phlebotomist to collect a blood sample. The questionnaire consists of questions from the VA's Million Veteran Program survey and other survey instruments chosen, in part, to permit comparison among other Gulf War Veteran population surveys. We conducted cognitive interviews of our developed survey instrument with Veterans who served during the first Gulf War Era to determine item comprehension, Veteran sensitivity issues, survey formatting and navigation issues. We will discuss respondent recruitment and selection, procedure sequences, observation recording and interpretation, recommendations made to the sponsor and resulting edits to the questionnaire. Through this presentation, we seek to encourage greater transparency in cognitive interviewing methods and results and provide insight into conducting cognitive interviews in Veteran populations.

## **5. Evaluation of Nonresponse Bias in a Cohort Study of World Trade Center Terrorist Attack Survivors**

Shengchao Yu, *New York City Department of Health and Mental Hygiene*  
Robert Brackbill, *New York City Department of Health and Mental Hygiene*  
Steven Stellman, *New York City Department of Health and Mental Hygiene*  
Sharon Ghuman, *New York City Department of Health and Mental Hygiene*  
Mark R. Farfel, *New York City Department of Health and Mental Hygiene*

The World Trade Center Health Registry (WTCHR) is a longitudinal study that tracks 71,000 enrollees to understand the long term health effects of the September 11 2001 terrorist attack. In 2003-04, 2006-07, and 2011-12, the WTCHR conducted three waves of health surveys. Response rates to the two adult follow-up surveys were 68% and 63% respectively. Falling response rates in follow-up surveys, especially among particular 9/11 populations at risk, pose critical challenges for evaluating associations between 9/11 exposure and health outcomes as well as prevalence estimates of health outcomes.

The objectives of this study are to assess: 1) nonresponse bias in the prevalence estimates of health outcomes among enrollees at follow-ups, and 2) the extent to which nonresponse may have affected estimates of association. We focus on two prevalent 9/11-related outcomes of public health importance, recurrent lower respiratory symptoms (LRS) and probable posttraumatic stress disorder (PTSD) as well as self-assessed general health.

We found that individuals who participated in all three surveys were less likely to report poor or fair health, probable PTSD, or LRS at follow-ups than those who participated in only one follow-up survey. However, such differentials were not present for other health outcomes including asthma, chronic bronchitis, gastroesophageal reflux disease, high cholesterol, hypertension, and anxiety.

We also found that attrition was not a substantial source of bias in the association between 9/11 disaster exposure and key physical and mental health outcomes. Specifically, the association between disaster exposure and probable PTSD and LRS was similar in magnitude and did not differ statistically between individuals who completed all three Registry surveys and those who participated in only one follow-up survey.

The absence of nonresponse bias in the association of disaster exposure and health outcomes thus far, despite modestly biased health estimates, provides a level of confidence in overall Registry findings.

## **6. Do You Really Love Your Neighbor?: A Comparative Analysis of Religious Tolerance in Two Democracies**

Austin Countryman, *University of Nebraska-Lincoln*  
Caitlin E. Deal, *University of Nebraska-Lincoln*  
Allan L. McCutcheon, *University of Nebraska-Lincoln*

Since 9/11, Muslims have become a more salient “outgroup” for many within mainstream American society, leading to diminished favorable opinions towards Muslims in America. Though religiosity in the U.S. is relatively high within Western nations, Islam is not widely represented in the population, and is seen as part of the religious and cultural outgroups (Kalkan, et al. 2009).



In June 2008, it was apparent that Barack Obama, misbelieved to be a Muslim by a significant proportion of the population, would be the Democratic Party's nominee for President of the United States (Pew Research Center, 2010). Obama's nomination increased the saliency of Muslims as part of the outgroup, increasing the likelihood that when asked to think about other religions, Americans may have been primed to think of their aversion towards Islam.

This paper seeks to evaluate religious tolerance in the United States, comparing it to the United Kingdom, another democracy with a similar culture, but lower levels of religiosity. Additionally, this paper examines the hypothesis that the U.S. exhibits lower levels of religious tolerance than the U.K.

Preliminary analyses of data collected in the summer of 2008 by the Gallup World Poll, a probability based sample, indicate that residents of the United States are more than twice as likely to consider religion important in their daily lives. Contrary to expectations, however, the analyses suggest that residents of the U.K. may be less tolerant towards other religions. Eighty-one percent of those surveyed in the U.S. strongly agreed that they would not object to a person of a different religious faith moving in next door, while only 57 percent of those surveyed in the U.K. said the same. The inquiry into the religious tolerance of these two nations allows us to examine the role of public opinion and democracy in comparative perspective.

## **7. Pairing a Qualitative Assessment and MaxDiff Surveys to Improve Gate Area Comfort at Seattle-Tacoma International Airport**

Mark E. Vande Kamp, *Port of Seattle*

Passenger service at airports can be improved most effectively by focusing efforts on service items that meet two criteria, a) the service is important to passengers, and b) the airport is not performing the service well. Accordingly, analyses of Sea-Tac passenger survey data suggested that there was an opportunity to improve customer service by addressing the service item, "Comfort of waiting/gate areas" (hereafter, "gate comfort"). However, because the service item was too general it wasn't clear how to improve gate comfort. In this research, complementary qualitative and quantitative methods served to identify and measure specific factors affecting comfort in gate areas. In the first, qualitative stage, open-ended interviews and passenger observations were conducted. These activities identified 15 general factors related to gate comfort. In the second stage, online surveys were conducted in which 219 respondents evaluated the importance of the 15 comfort factors and 244 respondents evaluated how well Sea-Tac was providing each factor. The surveys used the maximum difference scaling technique (MaxDiff) with results analyzed using hierarchical Bayes estimation. Importance and performance scores were calculated by transforming the weights to a ratio scale such that the item scores summed to 100. Plotting these scores on X and Y axes produced a quadrant analysis identifying items that had the most potential to improve gate comfort. Results suggested that managers should focus on three factors: 1) improvements to seating. In particular, more seats were desired; 2) sufficient space. Managers must balance the shortage of seats against passengers' desire for sufficient space to avoid feeling crowded by other passengers; and 3) support for personal electronics. Results related to electrical access and Wi-Fi illustrate how passengers increasingly depend on personal electronic devices. The qualitative and quantitative techniques came together in a complementary manner to identify these promising means of increasing passenger comfort.

## **8. Societal Inequality and Individual Well-Being: Results from 96 Societies and 300,000 Individuals, 1981-2008**

Jonathan Kelley, *International Survey Center and University of Nevada, Reno*

Mariah D. Evans, *University of Nevada, Reno*

Income inequality has been contentious for millenia, a source of political conflict for centuries, and is now widely feared as an undesirable accompaniment to economic progress. But equality is only a means to an end and so must be evaluated by its consequences. The fundamental question is Bacon's and Bentham's: Empirically what consequence does a nation's level of income inequality have for its citizens' well-being? We find that the answer depends on the society. In developing nations not formerly Communist, inequality is probably beneficial, certainly not harmful. In advanced nations it is irrelevant. Only for cohorts passing their formative years under Communism does inequality reduce happiness; for new generations in post-Communist societies it is irrelevant, or possibly beneficial. Data are from the pooled World Values Surveys and European Values Surveys, 247 representative national samples in 96 nations, 1981 to 2009, with over 330,000 respondents. Analysis is by variance-components multilevel models. These results are net of socioeconomic development (GDP) and a range of individual-level control variables. An important implication is that good measurement of individual subjective outcomes such as happiness and empirically assessing its determinants could lead policy formation in different directions from past practices of relying on elite assumptions about the determinants of happiness. A second important implication is the importance of measurement in situations where causal linkages are strongly context-dependent. For example, generalizations from advanced societies about the effects of inequality on happiness could lead a policy-maker in a developing country badly astray (and vice versa).

## **9. Capturing Minority Populations in Telephone Survey: Experiences from the Ohio Medicaid Assessment Survey Series**

Bo Lu, *College of Public Health, The Ohio State University*

Timothy Sahr, *Ohio Colleges of Medicine Government Resource Center*

Marcus Berzofsky, *RTI International*

Amy Ferketich, *The Ohio State University*

Jamie Ridenhour, *RTI International*

Rachel Tumin, *Ohio Colleges of Medicine Government Resource Center*

Often surveys are interested in oversampling certain minority populations in order to increase the precision of the estimates for those sub-populations. In a telephone survey this can be done with a landline frame by either targeting phone exchanges in certain Census tracts with higher concentrations of the sub-population of interest or by using listed samples of phone numbers with surnames in the population of interest (for ethnic targeting). However, as more individuals in these targeted sub-populations (e.g., young adults, African-Americans, Hispanics) move to cell phone-only phone use, landline-based oversample strategies become less effective. The 2012 Ohio Medicaid Assessment Survey (OMAS) oversampled ethnic minorities in Ohio using both approaches – African-Americans through Census tract targeting and Asians and Hispanics through list samples of surnames. The 2012 OMAS also uses dual-frame sampling design by including a substantial cell phone component. In this study, we describe the results of our experiences and offer possible suggestions on how to improve the efficiency of the oversample, considering the impact that

increased cell phone sampling may have on geographic targeted landline oversampling (e.g., metropolitan area African-American density sampling).

## **10. Does Political Party Affiliation Influence Views of Quality Among Those Who Interact with the U.S. Healthcare System?**

Kirstin W. Scott, *Harvard University*

Robert J. Blendon, *Harvard School of Public Health*

Support for the Affordable Care Act varies deeply by political party, with the majority of Democrats expressing favorable views of the law and its effect on healthcare quality and Republicans holding opposing views. Partisan differences have also been detected in aggregate polls when ascertaining public opinion on the healthcare quality and trust in healthcare professionals. It is not known, however, whether or not such differences persist among those who have had recent experiences with the U.S. healthcare system.

Does recent exposure to the healthcare system modify the relationship between political party and views on healthcare quality?

Data come from the “Sick in America” survey conducted by NPR/RWJF/HSPH in March 2012 among a nationally representative sample of 1508 respondents (MOE  $\pm$  3.1). Those who had been hospitalized in the last 12 months or had a serious illness were categorized as “ill” respondents (n=516, MOE  $\pm$  7.2). Subgroup analyses of ill versus total respondents, by political party affiliation, were done to compare views and experiences with healthcare quality.

More Democrats (38%) believe quality of U.S. healthcare is a major problem than Republicans (29%), however, ill Democrats (46%) and ill Republicans (41%) are comparable in their concern for quality. While more ill Democrats (90%) reported being satisfied with the quality of their care than ill Republicans (80%), no major substantive differences exist in the experiences with the healthcare system – and attitudes about physicians – among ill Democrats (n=281) and ill Republicans (n=145). Some differences, however, were identified in relation to medical testing, with more ill Democrats (25%) reporting that they had to wait for tests longer than ill Republicans (12%).

Though partisan differences exist on views on U.S. health reform and healthcare quality, substantive differences do not appear to persist after accounting for recent exposure to the healthcare system.

## **11. Public Perception Changes toward Real Estate Consumption in China: A Comparative Analysis of 2006 and 2009**

Mengyang Wang, *University of Nebraska-Lincoln*

Real estate is a critically important industry in the contemporary Chinese economic system. During the period from 2006 to 2009, land developments expanded with astonishing speed; many major cities have experienced sharply increased prices in real estate (Ahuja 2010). Population growth, combined with a fixed supply of land, has created an imbalanced development in the housing market, and driven housing prices higher (Chow 2009). However, there is little data on the changes in the public perception toward real estate consumption during this period. In this study, we conduct a comparative analysis based on

data in 2006 and 2009, collected by the Gallup World Poll, a multinational probability-based survey.

To examine whether the attitudes and perceptions of Chinese respondents regarding future housing purchases can be explained by the presence of an underlying construct related to the perception of real estate, simultaneous latent class analysis is used to examine the structure of perception patterns, as well as changes in patterns from 2006 to 2009. Specifically, this research focuses on three questions: (1) can Chinese respondents be categorized into distinct preference groups in terms of their planning for future purchases of real estate, (2) has the preference in housing purchases shifted (i.e. from dwelling to investing) between 2006 and 2009, and (3) has the structure of preference patterns changed over this period.

Preliminary analyses suggests a complex pattern of attitudes and perception toward real estate consumption among the Chinese people, providing evidence that the percentages of respondents with the intention of investing has remained the same (16%) even though the average housing prices increased 40% between 2006 and 2009. Meanwhile, the number of people who prefer to buy housing for residence decreased 8% during the same period.

## **12. Nonresponse Bias Analysis in a Genetics Sub-Study Among Reserve and National Guard Soldiers**

Christine Cowles, *Abt SRBI, Inc.*

Mark Morgan, *Abt SRBI, Inc.*

Researchers are increasingly combining biological and survey data to enhance understanding of the interactions between multiple biologic and behavioral factors associated with the risk of, or resilience to disease. A high level of participation in the biologic sample phase of studies is critical to the success of these research goals.

The Reserve and National Guard (RNG) study is a four-year longitudinal research project to assess mental health and service utilization among a national sample of Reserve and National Guard personnel. At baseline, 2,000 soldiers were surveyed by telephone. At the conclusion of the telephone survey in Wave 3, respondents were invited to provide saliva samples through an in-home self-collection method.

The research questions addressed in this analysis are: Does participation in the genetics study differ by meeting criteria for PTSD, depression, anxiety disorder or alcohol abuse? Which characteristics of military service are most strongly predictive of participation? Among those respondents who agreed to participate, are there differences between those who did and did not complete the biological sample collection?

Our initial review of the data indicates that 1,421 (71%) soldiers requested a saliva sample collection kit, and of those 41 withdrew from the genetics project. Of the 1,380 potential samples to be returned, 893 (65%) of the samples were mailed to the lab. Therefore we have a saliva sample from 44% of the RNG cohort and we will analyze whether there is nonresponse bias in the genetics portion of this study.

Finally, there will be a short qualitative analysis of the reasons given for declining the genetics portion of the RNG study to inform survey researchers about perceived barriers to participation in this type of biological sample collection.

### **13. The Impact of Using Cell Phone Zip Code Matching: The Case of Boston**

Marilyn Wilkinson, *Abt SRBI, Inc.*

Dennis Daly, *Abt SRBI, Inc.*

Barbara Fernandez, *Abt SRBI, Inc.*

Dan Dooley, *Boston Public Health Commission*

The Boston BRFSS defines residency in Boston using a very narrow range of zip codes. Starting in May 2013, in an attempt to improve sample efficiency and targeting, cell phone records were run through the ZIP match process to stratify by 'matched in Boston', 'matched outside of Boston', and 'unmatched'.

Our objective is to compare matching and rate center cell phone sample elements by major demographic and health indicators to determine: 1) differences in these populations, and 2) the impact the mix of these categories has on sample bias and representativeness.

To answer this question, we examine the 2013 Boston BRFSS cell phone sample cases by matching status, with major demographic variables and major health variables. Both cell contacts and completed interviews are assessed using Chi square significance tests.

Based on preliminary, unweighted data (May through September), one-quarter (26%) of the sample was matched with an eligible Boston zip code, an additional 13% was matched to an ineligible Boston zip code, and the remaining 61% was not able to be matched to a zip code. Among contacts, 23% of matched cases resulted in a complete as compared to 11% of unmatched cases and 4% of matched outside cases. We found significant demographic differences among these groups including cell phone status, age, ethnicity, race and income ( $p < .001$ ). Health outcomes examined included: asthma, smoking status, and diabetes. Higher rates were found when examining smoking status among the 'unmatched' group, although no differences were found for asthma and diabetes.

Our initial conclusions show that the 'unmatched' group accounts for a higher representation of the youngest age group, Hispanics, and non-whites, and therefore, needs to be included in a sufficient proportion to enhance representativeness. Improving efficiency in reaching cell phone users within narrowly defined geographic areas with sample manipulation needs to be balanced with reducing bias.

### **14. FIFA 2022: The Issue of Migrant Workers in the Arab Gulf Countries**

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

Kien Trung Le, *Social and Economic Survey Research Institute (SESRI)*

John Lee P. Holmes, *Social and Economic Survey Research Institute (SESRI)*

Sara Ali Ahmed Zikri, *Social and Economic Survey Research Institute (SESRI)*

Mohammed Hassan M.T. Al Subaey, *Social and Economic Survey Research Institute (SESRI)*

Semsai Al-Ali Mustafa, *Social and Economic Survey Research Institute (SESRI)*

When the Fédération Internationale de Football Association (FIFA) began its meeting last October about moving the Qatar 2022 World Cup to winter, the issue of treatment of migrant workers stretched a long shadow obscuring even the question of Qatar's searing summer sun. Human rights organizations have increasingly put pressure on Gulf Cooperation Council (GCC) countries to safeguard the rights of all migrant workers whose percentage of

the total population is so large as to constitute the majority of the labor force. Qatar exemplifies this issue since, according to the Qatar Statistics Authority labor survey, by 2010 the foreign labor force represented 94 percent of the Qatari population over 15 years of age. It is surprising therefore that despite the size of the foreign labor force in these countries, little is known about these blue-collar workers. An improved understanding of the lives and experiences of blue-collar workers in the Arab Gulf countries is essential to the development of labor force policies and planning in each of the GCC countries. It also has implications for the broader issue of the movement of workers and their concerns worldwide. Therefore, it is important to understand the demographic and social characteristics of blue-collar workers, the process by which they obtain employment, their remittances to their family members left behind, and the nature and frequency with which they encounter problems and challenges in their workplace countries. Accordingly, SESRI conducted the 2012 GCC Migrant Workers Survey, the first of its kind in the region, to provide an informational baseline of this population in the GCC. Respondents were asked about their attitudes, their living and working conditions, and their contact and communications with family members left behind. The results illustrate variation in blue-collar workers' challenges and difficulties across the GCC despite being under the same of immigration system.

### **15. Introducing Bias – An Evaluation of a Referendum Ballot**

Mikael Gilljam, *Department of Political Science, University of Gothenburg*  
Sebastian Lundmark, *University of Gothenburg*

This paper evaluates a referendum on congestion charges in the municipality of Gothenburg, Sweden, taking place in September 2014. The politicians in Gothenburg have decided to use an unbalanced question wording on the ballot, presenting an introductory text with their own justification of the congestion charges, as well as a question wording asking whether the municipality should “continue with the congestion charges”. Previous research has been in disagreement on the effects of balancing survey questions. Whilst Schuman and Presser (1981) and Narayan and Krosnick (1996) argue that balancing questions change the distribution of the answers, Shaeffer et al. (2005) argue that minimally balanced questions do not change this distribution. Perhaps the politicians sway the opinion by not providing a balanced or minimally balanced question wording.

This paper aims to, firstly, evaluate whether excluding the introductory text decreases the likelihood of voting yes in the referendum. Secondly, the project aims to test whether different question wordings change the likelihood of voting yes. Using a 2x2 full factorial experimental design on a probability sample of Gothenburg residents, comparing the ballot with or without the introductory text as well differences in wording (continue vs. abolish the congestion charges), this paper evaluates the ballot text in the referendum. Preliminary results indicate that providing positive information as well as using an unbalanced question wording biases the likelihood of voting yes on the referendum.

### **16. Echoes of Egypt: Understanding the Telephone-Owning Population in the Arab World's Largest Nation**

Samuel Solomon, *D3 Systems*  
David Peng, *D3 Systems*

In the last two and a half years, Egypt has experienced a dizzying series of political transformations: a dictator's fall, military rule, an Islamist presidential administration, and a popular military coup. As the most populous state in a Middle East undergoing deep and

lasting changes, understanding the views of the Egyptian populace is now more important than ever.

However, at a time when public opinion plays an increasing role in the politics of Egypt, chronic instability presents significant challenges to face-to-face surveys of public opinion. As a cost-effective means to bypass these challenges and directly reach respondents in-country, telephone surveys are the likely future of opinion research in Egypt. It is therefore critical to examine who composes Egypt's telephone-owning population so researchers understand the biases implicit in sampling Egypt's national population via telephone. The authors of this poster make use of a nationally representative face-to-face survey to explore the demographics and telephony habits of Egypt's telephone-owning population and better inform future telephone-based public opinion research of Egypt.

This poster utilizes data collected in Egypt during October and November of 2013. D3 Systems, Inc. commissioned this national face-to-face survey using a stratified multi-stage cluster design. Sampling procedures included the enumeration of households at sampling points and subsequent selection of households for inclusion in the survey. The poster includes details about the survey's methodology and challenges faced by the researchers while conducting this project.

#### **17. Examination of Order Effects on the Presidential Candidate Trait items of the ANES and the Importance of Context: Partisanship, Trait and Candidate**

Vasiliki-Maria Agalioti-Sgompou, *Institute for Social and Economic Research*

Although it has been long established that the order in which questions are listed in a questionnaire affects the answers given by the respondents, the contextual factors that influence the response have not been observed sufficiently (Tourangeau, Rips and Rasinski, 2000). The present research focuses on the order effect in the Presidential Candidate Traits from the 2008 American National Election Studies. In the Trait questions, respondents are required to evaluate how well each trait/characteristic (for example, moral, honest, strong leader) describes each presidential candidate. The order in which each respondent has assessed the Republican and the Democrat candidate has been randomized. In accordance with existing research on contrast/assimilation effects (Bless and Schwarz, 2010; Herr, 1986; Herr, Sherman and Fazio, 1983), our findings show that respondents with different strength in partisanship identity are affected differently by the order of the questions. More specifically, they are more likely to assess negatively the candidate of the opposite political party if he is asked first (for example, a strong Democrat for McCain). Moreover, we observe in our findings certain contextual factors interacting with the order effect, especially the candidate and the trait assessed, which can be associated with circumstantial intricacies in the formation of public opinion. We conclude with a discussion on cognitive and emotional process behind survey response.

#### **18. From Bully to the Boss: Governor Chris Christie's Post-Sandy Stardom and the 2013 New Jersey Gubernatorial Election**

Ashley A. Koning, *Rutgers University*

David P. Redlawsk, *Rutgers University*

From the moment he took office as governor of New Jersey in January 2010, Chris Christie became a polarizing figure both state and nationwide, loathed by Democrats for his battles with teachers' unions and state workers yet loved by Republicans for his no-nonsense

rhetoric, fiscal conservatism, and potential for the party's fundraising efforts and future. Most of 2012 proved contentious for Christie as speculation over a presidential bid abounded and as New Jerseyans began to desire new leadership. But the arrival of Hurricane Sandy changed both the physical and political landscape of the state, rocketing Christie to bipartisan stardom and creating a prolonged "Sandy" effect that made his victory in New Jersey's 2013 gubernatorial race inevitable. We explore Christie's trajectory from bully to bipartisan hero throughout his first term, reelection campaign, and beginning of second term. We particularly focus on the effects Hurricane Sandy had on Christie's reelection prospects, tracing the importance of Christie's job approval in various issue areas – including Sandy recovery – on New Jersey voters' decision for governor. In an analysis of the New Jersey governor's race from Sandy through Election Day, we find that Christie's large lead was primarily sustained by a Sandy momentum – rarely if ever seen in polling post-disasters or other "rally 'round the flag" moments – despite lower approval ratings on problems voters in the state rank as more important. Christie's high favorability ratings and reelection have thus been the result of Sandy-induced affect among New Jersey voters – inspired more by emotion than cognition. Post-storm, voters now consistently see Christie as a strong leader, and previous liabilities, like his tough-talking attitude, have now become assets as he enters his second term – and embarks upon a possible presidential bid for 2016.

## **19. A Nation Divided: Polling Results and the Political Division of the American Public**

Evan Welty, *Northern Arizona University*

It is clear that the American electorate is changing, but how much? I seek to divide the overall political opinions of Americans into 5 categories: far left, center left, center, center right, and far right. These categories will be based on, and applied to contemporary political issues of: military intervention, civil liberties (domestic surveillance, abortion, gay marriage, and marijuana legalization), immigration reform (with allowing current undocumented persons to stay in the country), the Affordable Care Act, environmental protection, and regulations, Keystone XL, among others. Additionally, I examine conspiracy theories (9-11 truthers, birtherism, viewing President Obama as a Muslim, concern about the Benghazi terrorist attack, ext), and trends (Vegan dietary rates, ext). Polling on these issues is then compiled to come up with an accurate snap shot of the political make up of the country. This polling shows overlap in a few key areas, and over all shows a substantial polarization between the center left and far right, which is where most people in the United States are currently politically. However, it also shows substantial disconnect in many areas from the policies of the United States government. Lastly, this information is compared against voter participation rates per group to explain recent election results.

## **20. Perceptions of Marcellus Shale (MS) Drilling as Economic Opportunity, Environmental Threat, or Both? Socio-Demographic Profiles and Relationships with Environmental Attitudes/Behaviors in the General Population and Among Residents of a High MS Activity**

Scott Beach, *University of Pittsburgh, University Center for Social & Urban Research*

Janet Schlarb, *University of Pittsburgh*

Bernard Goldstein, *University of Pittsburgh*

The significant increase in Marcellus Shale (MS) "hydrofracking" to extract natural gas has led to public debate about economic opportunities relative to potential environmental threats.



Understanding public attitudes towards MS activity helps to inform the debate. This paper presents data from two landline / cell RDD telephone surveys, with several parallel survey items: (1) A general population survey of residents of the 7-county Pittsburgh MSA focusing on environmental attitudes / behaviors (Spring 2013; n=805); (2) a general population quality of life survey (including a section on MS and the environment) of adults residing in Washington County, PA; a high MS-drilling area south of Pittsburgh (Fall, 2011; n=502). In both surveys, respondents were asked to rate MS drilling as an economic opportunity and a potential environmental threat (none, only a little, moderate, significant). Three groups are examined: (1) respondents perceiving a moderate or significant environmental threat and no or only a little economic opportunity (anti-MS environmentalists; 16%); (2) respondents perceiving a moderate or significant economic opportunity and no or only a little environmental threat (pro-MS economic supporters; 39%); and (3) respondents perceiving both a moderate or significant environmental threat and economic opportunity (i.e., complex / ambivalent MS attitudes; 45%). The paper presents comparisons of the three groups on socio-demographic and environmental attitudes / behaviors (including other MS attitudes) drawing from both surveys. Detailed comparisons are made between sub-groups of complex / ambivalent respondents who support MS activity overall (about 35%), those who oppose MS activity (35%), and those who neither support nor oppose MS (30%). Preliminary results show distinct socio-demographic profiles; and that the complex / ambivalent group is similar to environmentalists on some environmental variables, similar to economic supporters on others, and distinct on others. MS attitude patterns also vary somewhat between the general and high MS activity location populations.

## **21. Change or Confirmation? Evaluating the Impact of Superstorm Sandy and Public Views on the Existence of Global Warming**

Christopher P. Borick, *Muhlenberg College Institute of Public Opinion*  
Barry G. Rabe, *University of Michigan*

When Superstorm Sandy slammed into the east coast in late October of 2013 the public discourse in the United States regularly included discussion of the role that climate change may have played in the development and strength of the storm. While the scientific community continues to examine the role that climate change may have played in terms of this powerful storm, public opinion researchers have become engaged in measuring the effects that the storm may have had on American attitudes and beliefs regarding the issue of global warming. Did the storm and its devastating impact on the northeast cause any shifts in public acceptance of global warming. Using data from the National Survey on Energy and the Environment (NSEE) that was gathered shortly before and after Sandy's landfall, including over 300 interviews with the same individuals, the storm's effect on public opinion regarding climate change can be measured. The findings suggest that the storm had very limited impact on overall beliefs regarding the existence of global warming among Americans, but instead was used as a source of confirmation for those who already accepted that global warming was happening. Conversely, among those who were skeptical of global warming's existence, there was a shift away from explanations that involved weather observations to more political and religious oriented factors.

## **22. Development of a Survey Tool to Assess Climate on Campuses**

Sowmya Anand, *Survey Research Laboratory*  
Linda Owens, *Survey Research Laboratory*

Campus climate can be defined as “the current attitudes, behaviors, and standards and practices of employees and students of an institution” (Rankin & Reason, 2008). Educational institutions often conduct assessments of the climate on their campuses as these are thought to affect academic outcomes for students (Hurtado, Griffin, Arellano, & Cuellar, 2008). While many studies primarily focus on diversity and inclusion aspects of climate (Hurtado et al., 2008), others adopt a broader definition, and include aspects such as university policies and research and scholarship (Rankin & Reason, 2008). Using this latter broader definition of climate, we seek to understand the dimensions underlying perceptions of climate and the extent to which they influence overall perceptions of climate. In our study, we started by compiling a list of questions used by various educational institutions to assess climate. After categorizing these 500 or so questions into various themes, we selected items from each theme to construct an assessment tool, after having made wording or format changes indicated by questionnaire design principles. Using the resulting instrument, we collected data via online surveys from faculty, staff, and students at two large universities in the Midwest. On the basis of exploratory principal components analysis and confirmatory factor analysis, we define dimensions underlying campus climate, which are different for employees and students. We then construct scales for these dimensions and establish their internal consistency. Next, via regression procedures, we establish the convergent and discriminant validity profiles for each. Finally, we present the results of regression analyses to understand the relative influence of the dimensions of climate on overall perceptions of climate. We assess these final dimensions in light of Rankin and Reason's typology of aspects that influence campus climate.

### **23. The Cannabis Debate in Mexico: Prohibition or Legalization?**

Diana Paola Penagos Vasquez, *Parametria SA de CV*

Francisco Abundis Luna, *Parametria SA de CV*

Jose Alberto Vera, *Parametria SA de CV*

Data from the most recent drug report of UNODC shows that cannabis is the most consumed substance in the world, and the United States is one of the main consumers. Besides this, the most recent National Survey on Addictions in Mexico found that the most commonly used illicit drug is marijuana.

Unquestionably drug trade and drug legalization are themes which deserves to be treated carefully. The proximity to the largest consumer of marijuana in the world, requires the Mexican government to legislate about it. We can not forget that American public trends favor the legalization of drugs. Also, due to the multidimensionality of this topic, it is important to note that the positions in civil society are diverse. In order to know the Mexican public's opinion on the issue of drug legalization, Parametría performs the following research since 2008.

The violence generated by the drug war has widespread the social perception that this fight is being lost. This result is consistent with recent data from the Pew Research Institute. The results show that 8 of each 10 Mexicans tend to stand against the legalization of drugs. Among all drugs, marijuana is of the lowest social rejection. Although the Mexicans do not favor the legalization of marijuana, they do approve it for medicinal use.

The road to be traveled in this debate is very long. Mexico should prepare its policy and market, to the gradual decriminalization that is currently developing in its main consumer, the United States. The failure of the punitive prohibitionist strategy against the trafficking of

illicit drugs, lead us to expect an increased production and consumption worse than in the beginning. It is necessary to create useful tools to combat drug use; these must be accompanied by a more flexible perspective

## **24. Civil Liberties and Security: The Public's Contrasting Values and Policy Preferences**

Jennifer Benz, *AP-NORC Center for Public Affairs Research*  
Daniel Malato, *AP-NORC Center for Public Affairs Research*  
Trevor Tompson, *AP-NORC Center for Public Affairs Research*  
Jennifer Agiesta, *The Associated Press*  
Becky Reimer, *AP-NORC Center for Public Affairs Research*  
Dennis Junius, *The Associated Press*

The terrorist attacks of September 11, 2001 were a defining moment for a generation. The years that followed saw significant changes in government policy regarding security, privacy and civil rights. Recent events cast the debate over increased security measures encroaching on civil liberties and privacy concerns in a new light. The American public continues to grapple with these issues. As a follow-up to a 2011 AP-NORC Center survey on the same topic, we conducted a national survey of 1,008 adults in the United States in August 2013. This polling shows opinions and preferences that sometimes appear to be in direct conflict with one another. Where do Americans draw the line when it comes to the monitoring of their communications and other actions taken by the government to keep them safe? Why do some Americans who self-describe as conservative end up holding more liberal views on many of these policies? Answering these questions and comparing the results of the 2013 survey to the 2011 survey provides a useful picture of how attitudes toward civil liberties and security policy continue to evolve in America. The findings suggest that Americans value both civil liberties and security and may not see some of the conflicts between policies designed to protect each one. Americans are more likely to say it's more important for the government to protect the rights and freedoms of citizens than to keep the public safe from terrorism, but are also more likely to say it is sometimes necessary to sacrifice rights and freedoms in order to prevent terrorist attacks. On specific policies, there is modest support for some seemingly-invasive proposals to prevent terrorist attacks on the one hand but complaints about the government's track record of protecting rights and freedoms on the other.

## **25. Domestic Hyperpartisanship and Foreign Policy Attitudes**

Gregory Holyk, *Langer Research Associates*  
Dina Smeltz, *Chicago Council on Global Affairs*

We live in hyperpartisan times. Congress is dysfunctional. A budget hasn't been passed since 2009. The recent stalemate over the budget and the government shutdown are just the most recent illustrations that cooperation and compromise in the Beltway are difficult to find. This high level of partisanship among political elites is both a reflection of and influence upon greater partisanship in the American public. Approval of the current president and his predecessor has shown the largest partisan gaps in history.

But what about attitudes toward foreign policy? Does politics really end at the water's edge? Is the recent hyperpartisanship on domestic issues reflected in foreign policy views? In this paper we compare Republicans and Democrats across 40 years of Chicago Council on Global Affairs surveys dating back to 1974. Across a wide range of issues, including support

for the use of troops, support for engagement, preference for multilateralism, foreign policy goals and perceptions of threats, views on China, immigration and climate change, among others, there is little evidence of greater distance between partisans today.

We also compare red and blue districts on these same issues in the latest Chicago Council survey from 2012. Again, though Americans from red and blue districts differ in degree on some issues, the difference is rarely substantial. On average, for all 85 questions included, the difference between them was just 3 percentage points.

One of the most interesting shifts in partisan politics has been the growth of independents and how their views differ from Democrats and Republicans. Our paper will demonstrate that it's independents who are more isolationist in general, less "hawkish" than Republicans on specific issues and less "dovish" than Democrats on others. And they've become more so of late.

## **26. What School do I Choose?: Determining the Accuracy of Self-Reported Data with Administrative Data**

Patrick Stark, *American Institutes for Research*

Jeremy Redford, *American Institutes for Research*

A key component to successful survey research is data accuracy, which can be identified by comparing self-reported data to administrative data. Inaccurate self-reported data may be a result of a lack of information on the part of the respondent. In many cases it is hard to determine the validity and reliability of self-reported data, because verifiable information is unavailable. With the growing debate over using school choice as an alternative to traditional schooling in the United States, there is a need to investigate to what extent parents understand the types of schools that their children attend, and the accuracy of self-reported data about school choice. This is particularly pertinent to education surveys that ask about school choice, as it is an important component of social and educational research.

Using the Parent and Family Involvement Survey (PFI) from the 2012 National Household Education Surveys Program (NHES:2012), we investigate the accuracy between parent-reported data about children's school type (i.e., Public, assigned; Public, chosen; Private, religious; and Private, non-religious) and administrative data about school type from the Common Core of Data (CCD) and the Private School Universe Survey (PSS). NHES:2012 is a nationally representative survey of school-age children in the United States that asks questions related to school choice and parent involvement in education and links CCD and PSS data to the focal child.

The research questions we intend to explore are: 1) What percentage of parent-reported data differ from school-reported data from the CCD/PSS? 2) Do these differences vary by school, and student and family characteristics? Initial findings indicate that in approximately 5 percent of cases parent-reported data differ from CCD and PSS data across different school types. The implications regarding school choice and the accuracy of self-reported data compared to administrative data in education surveys will be addressed in this discussion.

## **27. Evaluation of Nonresponse Bias in a Survey of Public Trust towards Official Statistics and Attitudes Towards Use of Administrative Records**

Rajesh Srinivasan, *Gallup, Inc.*

Manas Chattopadhyay, *Gallup, Inc.*

Morgan Earp, *U.S. Bureau of Labor Statistics*

Ryan W. King, *U.S. Census Bureau*

On behalf of the U.S. Census Bureau, Gallup conducts a longitudinal telephone study to better understand America's knowledge of and attitudes towards the FSS (Federal Statistical Systems) as a means to improve the agency's overall survey response rates and data quality. This paper reports the findings of a follow-up nonresponse bias study that was conducted in 2012 to examine the effect of nonresponse on the results of the main study. Data from those who participated in the main study and those responding to the follow-up nonresponse study were compared. A commonly used method to adjust for nonresponse is weighting of survey data by weighting adjustment cells based on selected socio-demographic variables. In addition to the standard method, we will also be constructing a response propensity model using auxiliary variables and data from the nonresponse study. Estimates of key survey measures using the two different weighting schemes will be compared to evaluate the extent of nonresponse bias.

## **28. Predictors of Response Rate among Poor Populations of Developing Countries: Results from the Baseline Evaluation of Salud Mesoamérica 2015**

Ali H. Mokdad, *Institute for Health Metrics and Evaluation, University of Washington*

Marielle C. Gagnier, *Institute for Health Metrics and Evaluation, University of Washington*

K. Ellicott Colson, *Institute for Health Metrics and Evaluation, University of Washington*

Erin Palmisano, *Institute for Health Metrics and Evaluation, University of Washington*

Alexandra Schaefer, *Institute for Health Metrics and Evaluation, University of Washington*

Brent Anderson, *Institute for Health Metrics and Evaluation, University of Washington*

Annie Haakenstad, *Institute for Health Metrics and Evaluation, University of Washington*

Bernardo Hernández Prado, *Institute for Health Metrics and Evaluation, University of Washington*

Rafael P. Lozano, *Institute for Health Metrics and Evaluation, University of Washington*

Response rates in health surveys are declining globally and are affected by many factors. We sought to examine individual, household and community characteristics associated with variation in unit and item response rates using data from the Salud Mesoamérica 2015 Initiative in six countries.

In each country, we randomly selected localities with probability proportional to size and approached all households with a brief survey of demographic characteristics. Administration of this survey improves the targeting of eligible households, response rates, and adjustment for nonresponse bias. Randomly-selected households with eligible women or children were asked to participate in a comprehensive second survey containing sensitive questions on topics including family planning behaviors, reproductive history, child anthropometric measures and anemia testing. All data was collected with real-time computer-assisted personal interviews administered in common local languages. This evaluation compares unit and item response rates between countries and by community- and individual-level characteristics.

Country-level unit response rates for the first survey ranges from 97.7% in Mexico to 99.3% in Panama. The second survey has lower unit response rates of 90.4% in Guatemala to 97.8% in Nicaragua. While participants' unit response rate is higher in Panama compared to Mexico, item response rates are lower. For example, respondents in Panama were less likely to respond to questions on contraceptive use. Differential item nonresponse is observed in Panama for child anemia testing between children whose mothers never attended school (21.1% of children received consent for blood tests) and those who attained primary (54.0%) or secondary or higher (70.5%) education. There is no variation in blood testing item nonresponse by maternal education in any other country.

These patterns in participation indicate the need for implementation of different strategies to improving response rate in low-resource settings of developing countries. Moreover, our findings enable researchers to design post-stratification adjustment of weights in surveys.

## **29. Voices of the Democracy Over 75 Years**

Lois Timms-Ferrara, *Roper Center for Public Opinion Research*

Robert Petrin, *Ipsos*

Chris Jackson, *Ipsos*

David Kolarik, *Ipsos*

The Smithsonian Institution thoroughly and creatively documents the history of America's scientific inventions, art and culture. The National Archives preserves the original written words that declared the values of the nation's founders and established the form of government that is still globally revered. Individual presidential libraries chronicle every decision made by each administration. Suffice to say the Library of Congress has left few actions of our law making bodies undocumented. Indeed, it is difficult to imagine the nation without these resources that have so scrupulously detailed our rich heritage.

Not only the words and works of the leadership of this great nation have been accounted for, but also the values, opinions, and attitudes of its populace. Since the 1930s, with the advent of modern, scientific polling, public opinion data has been archived in a unique research center. Pioneer pollster, Elmo Roper foresaw the value of protecting those early data when he founded the Roper Center public opinion archives in the 1940s, "A successful democracy is dependent upon public opinion."

This distinctive collection has grown exponentially over the last five decades.

This poster will demonstrate the breadth of the polling collections archived at the Roper Center, documenting the voices of the people over more than seven decades. The value of this large body of opinion data lies in how researchers are able to study issues across survey houses, question wordings, time periods and methodologies. Topically, these collections trace views of the US public on everything from World War II to the civil rights movement, to 21st century foreign policy. From social issues to consumer matters, this extensive compilation covers everything from opinions on specific policies to broad value-driven beliefs.

### **30. Changing Horses (or Sources) Midstream: Comparing Variance in Different Sample Sources Over Time in a Blended Sample Online Survey**

Zachary Lewis, *Ipsos*  
Robert Petrin, *Ipsos*  
Chris Jackson, *Ipsos*  
David Kolarik, *Ipsos*

Online surveys, using both opt-in panels and river sampling (one-time recruitment of potential survey respondents while they are engaged in other online activities), are becoming an increasingly common method of data capture in an era when survey costs are rising and response rates from more traditional survey modes are declining. In spite of the promise of these methods to address important challenges facing the survey research community, little is known about how river samples perform when used as a supplement to traditional panel. This paper will seek to address this shortcoming by comparing the variance in responses to questions addressing specific economic issues from Ipsos panel respondents to the variance in responses to the same questions from Ipsos river sampling respondents. Comparisons will be made by key demographic and geographic characteristics, and responses will be benchmarked against key economic indicators to derive an approximation of the survey error introduced by river sampling.

### **31. Seasoned vs. Fresh: A Comparison of Political Opinion Survey Panel Respondents**

Catherine Wilson, *23andMe*  
Matthew DeBell, *Stanford University*

As online panels become more widely used, survey respondents become more likely to have experience completing previous surveys. In some cases, they may also have a financial motivation to complete many more surveys. In this paper we study how respondents in a panel may differ from each other based upon the number of surveys they have previously completed. We use American National Election Studies (ANES) data from the Evaluations of Government and Society Study (EGSS) to investigate differences between seasoned respondents who have completed many surveys (sometimes even more than a few hundred surveys) and fresh respondents who have completed only a handful of surveys. We show how seasoned and fresh respondents are similar in many respects and also different in a number of respects including on both demographic and attitudinal dimensions. We show how seasoned respondents may impact the conclusions of a survey research project, and suggest ways in which this may be controlled.

### **32. Cognitive Interviewing to Explore Perceived Question Sensitivity in the Kingdom of Saudi Arabia**

Kristen L. Cibelli, *Program in Survey Methodology, University of Michigan*  
Zeina Mneimneh, *University of Michigan*  
Mona Shahab, *Prince Salman Center for Disability Research*  
Yasmin Altwaijri, *King Faisal Specialist Hospital and Research Center*

Several studies have examined the use of cognitive interviewing techniques among speakers of non-English languages in the United States and Europe. However, little attention has been paid to the local application of this method in other cultural contexts. This paper examines the application of cognitive interviewing in the Kingdom of Saudi Arabia (KSA) to elicit problems with questions, particularly on items of a sensitive nature. The local

research team conducted the cognitive interviews (n=65) in collaboration with the Survey Research Operations at the University of Michigan in preparation for the National Health and Stress Survey (NHSS) in KSA. This survey is part of the World Mental Health Initiative. The cognitive interviewers used mainly scripted probes to explore possible problems with comprehension, recall and sensitivity for many survey items.

Question sensitivity is a key concern in the KSA context where mental health problems (the main focus of the interview topics) are culturally taboo. Unfortunately, little guidance exists in the literature on appropriate probes and probing techniques to elicit respondent feedback on question sensitivity. This study rotated several different probes to explore perceived sensitivity of specific survey items. Probes included “How difficult was it to talk about this issue?”; “Do you think that others would find it uncomfortable to talk about this issue?”; and “Tell me more about your opinion on this question”. We compare the effectiveness of these different probes in eliciting feedback on question sensitivity.

We also analyze how the cognitive interviewing results vary in terms of codable responses, number of problems elicited, and level of detail by respondent attributes (e.g. age, sex, income, education), interviewers and respondent answers to a debriefing questionnaire (e.g. difficulty thinking aloud, and discomfort with the process). We conclude with future directions for cognitive interviewing research related to sensitive questions and its application in cross-cultural contexts.

### **33. Understanding Consent for Physical Measurements, Biomarker Collection and Administrative Data Linkage in the Health and Retirement Study**

Colleen McClain, *Michigan Program in Survey Methodology*

Sunghye Lee, *Michigan Program in Survey Methodology*

Jessica Faul, *University of Michigan Research Center*

The wealth of information that can be collected via contemporary surveys, including interview data, biomarker data, and consent for linkage to administrative records, provides intriguing avenues for the analysis of survey data. However, data collection beyond interview questions complicates administration as the collection of physical measures and biomarkers, as well as administrative record linkage, requires additional consent from survey respondents. Consequently, complicated errors can be introduced on such measures when the obtained data are conditional on receipt of consent. In particular, consent for some measures but not others may lead to differential nonresponse errors.

The Health and Retirement Study (HRS), a longitudinal study of the elderly population, has collected physical measures and biomarker data (saliva and blood spots) since 2006 without engaging special examination facilities. Additionally, HRS solicits permission from respondents to link their survey data to the Social Security Administration (SSA) records on earnings and benefits. The study provides a unique opportunity to examine consent patterns to these requests collectively and separately in cross-sectional and longitudinal fashions.

Using HRS data, we examine overall patterns of three different sets of consents: those for physical measurements, biomarker, and SSA records. Specifically, we attempt to ascertain whether consent rates are related to respondents' socio-demographic characteristics (e.g., age, gender, race/ethnicity, interview language, education, income, employment status) and physical, mental and cognitive health status, their resistance to the current survey request, and their cooperation to future data collection, as well as whether the patterns differ across



measures. We also incorporate detailed interviewer observations and interviewer characteristics into the analysis in order to better understand the consent patterns. In doing so, we particularly focus on the cultural background of respondents, including race, ethnicity and interview language, as well as that of interviewers.

### **34. Drowning In a Sea of Regulation: Perception or Reality and the Policy Implications of Finding the Truth**

Debbie Borie-Holtz, *Rutgers University*

Stuart Shapiro, *Rutgers University*

Congress is currently considering a dozen regulatory reform bills to modify the procedures by which executive branch agencies issue new regulations. Supporters of these bills hope they will increase the economic efficiency of regulations and bring more accountability to the regulatory process. Opponents argue they will impede agencies' ability to issue new rules and compromise public health.

In *The Politics of Regulatory Reform*, we argue that both supporters and opponents overstate the impact of procedural reforms on regulatory policymaking. Instead of seeking to determine who is right, we should be asking why, given limited evidence of regulatory reforms' effectiveness, politicians continue to be fascinated with these policy initiatives.

We believe the answer to this question stems from a public perception that rulemaking is harming businesses and states' economies. Although the actual effects of regulation are at best unclear, it's very clear that business executives believe that regulations hinder their profit margins.

In surveying executives in five Midwestern states, we found widespread concern about regulation inhibiting business activity. It matters little whether this perception is the result of a genuine phenomenon not picked up in the economic literature or the result of political rhetoric that demonizes regulation. To quell constituents' perceptions about this threat, elected officials find it politically efficacious to implement regulatory reforms.

What follows is a cycle in which politicians respond to a perceived problem, absent the empirical evidence that one exists. Politicians reaffirm that regulations are harmful by sponsoring reform laws, even if such laws are ineffective. In so doing, policymakers reinforce the perception that regulations hurt businesses.

We return to these states – one year later in January 2014 -- to measure attitudes about regulations. We hypothesize that perceptions about “job killing regulations” will have diminished as local economies have improved; supporting our claims of inefficient regulatory policymaking.

### **35. Giving It Your All: How Response Formats Affect Measures of Donation to Causes**

Frances M. Barlas, *GfK Custom Research*

Randall K. Thomas, *GfK Custom Research*

One measure of people's engagement with various political and social causes is the extent to which they donate money. To measure donations to various targets, researchers face a choice of which response format to use. The Multiple Response Format (MRF) is a

frequently used response format where people are presented with a series of elements and asked to select all of the targets to which they contributed. Another response format option is the Yes-No Grid (YNG) where respondents are asked to respond with either a 'Yes' or 'No' response to each cause. Prior research has indicated that the YNG tends to yield higher levels of behavioral self-report than the MRF. We conducted a web-based survey with opt-in panelists, asking whether they contributed to: a political candidate; a political party; an organization involved in political issues; a non-profit non-political organization; or a religious organization. In comparing answers to both response formats, we found a similar ordering of the targets in their rates of endorsement (i.e., most and least endorsed). However, there was a significant difference between the response formats in the rates of self-reported contributions to the various targets with the YNG having significantly higher self-reports of contributions across all targets. We further examined how much money respondents contributed to the causes to which they reported having donated and found significant differences in the self-reported average amount contributed between respondents answering the YNG and the MRF. The additional responses that the YNG yielded may be due to respondents endorsing less salient targets.

### **36. Interviewer's Arguments to Persuade Potential CATI Respondents**

Yfke Ongena, *University of Groningen*

Marieke Haan, *University of Groningen*

Interviewer contact attempts are usually more effective than contact attempts without live interaction like mail or email, due to the possibility of maintaining interaction and tailoring argumentation. However, we don't know which arguments interviewers can best use in this interaction. Therefore, we designed an experiment in a CATI survey to study the effects of different arguments. Two groups of interviewers were instructed differently with respect to the set of arguments they were allowed to use. The most striking result is that in most cases that end up in an interview, respondents comply to be interviewed immediately after the interviewer's request. Thus, in those cases additional interviewer argumentation did not play a role for respondents to make their initial decision. Focusing on respondents that did not comply immediately and the argumentation of the interviewer in these cases, our results show that it is useful to give any argument. The response rate is almost three times higher when interviewers provide an argument (28%) than when they do not provide an argument (8%) after the respondent's initial refusal move. This effect is the same in both experimental groups. The most effective persuading strategies appear to be mentioning something about the duration of the interview, stressing importance of the survey, mentioning the interest in the respondent's age, and suggesting to at least trying to complete the survey. A significant difference in response rate between the experimental groups was found for interactions during which respondents ask about the survey institution; in this case the response rate is much higher for interviewers who were not allowed to explicitly refer to the university than for interviewers who were instructed to explicitly refer to the university. Therefore, we assume that more spontaneous reactions of interviewers are more effective than reactions they were especially trained on.

**Friday, May 16**  
**3:15 p.m. – 4:15 p.m.**  
**Demonstration Session 2**

**The Impact of Methodological Changes on the Comparability of Estimates Across Year in the California Health Information Survey**

YuChing Yang, *UCLA Center for Health Policy Research*

Hongjian Yu, *UCLA Center for Health Policy Research*

Ninez A. Ponce, *UCLA Center for Health Policy Research*

Alterations in methodology are inevitable for every ongoing population survey in order to accommodate social and demographic changes in the population being surveyed. In this study, we focus on two primary methodological changes that were implemented in the most recent 2011-2012 cycle of the California Health Interview Survey (CHIS) and the impact of these changes on the comparability of estimates from this cycle to those from previous CHIS cycles. First, CHIS adopted a new sampling design and weighting process to account for the proliferation of cell phone only households. Second, CHIS began using the 2010 decennial Census enumeration data as the base for its weighting control totals. The previous CHIS cycles used 2000 Census data. This transition to the new Census data affected both the weighted population size as well as demographic composition of the population.

We examine the impact of these two methodological changes on the comparability of estimates across years, using three methods for adjusting the estimates. Each adjustment “controlled” for a different aspect of the changes in survey methodology. The results suggest that some of the shifts in estimates between years can be attributed to the changes in methodology rather than changes in the population, so users should be cautious when trending estimates over this period. We propose that researchers interested in monitoring changes over time adopt standardization techniques that are commonly used in epidemiology to adjust their estimates.

**Roper Center: Archiving Services and Access Tools**

Marc Maynard, *Roper Center for Public Opinion Research*

Lois Timms-Ferrara, *Roper Center for Public Opinion Research*

Founded at about the same time as AAPOR, the Roper Center archives are now the largest and most comprehensive archives of public opinion data. The Center’s role in the data life cycle is one of preserving survey data entrusted to its care in perpetuity and making these data available via intuitive access tools.

In summer 2013, the Roper Center released a set of enhanced services impacting access to some 20,000 US and international survey datasets archived at the Center, as well as, iPOLL a database of more than 600,000 US questions and responses asked over the last 75 years. This demonstration will review recently improved data discovery and analysis tools that support the utilization of public opinion surveys. Survey practitioners engaged in questionnaire design, comparative research, and analysis of all types of survey data will discover the value of this collection unlocked by these advanced features.

**Friday, May 16**  
**4:15 p.m. – 5:45 p.m.**  
**Concurrent Session F**

## **Small Area Estimation**

### **Review of Methodological and Practical Developments in Small-Area Estimation**

Stanislav Kolenikov, *Abt SRBI*

Small area estimation (SAE) is an area of active growth in survey statistics that emerged in the 1990s following greater availability of computing power. It concerns the problem of obtaining reasonable estimates for domains where small sample sizes do not allow direct estimation using survey data only (including domains with zero sample observations), e.g., at the levels of a county or a metro area. Since typical national surveys contain at most several dozen observations at these levels of geography, models have been developed to support inference by borrowing strength from across the whole data set. Examples of SAE systems include SAIPE and SAHIE programs at the Census Bureau; Small Area Estimates for Cancer Risk Factors and Screening Behaviors at the National Cancer Institute; and statewide and large metro area estimates of prevalence of cell-phone-only households and persons based on NHIS.

The modern approach to small area estimation involves use of statistical models to predict the variable of interest. A regression-type model with random effects of small areas is fit to the data. Predictions from the model (synthetic estimators) are obtained and are then combined with the direct estimators based on data from the small area to produce composite estimators with reduced mean-squared error. A different approach applies the models fit on a special-purpose data set to a larger, high coverage data set, such as ACS, to produce model-only estimates.

This talk will present the main extant models for continuous and binary response, and discuss challenges in finding good explanatory variables, definitions of local geography, measures of accuracy and precision, and software packages to fit small area models.

### **Comparison of Small Area Estimation Methods for Use by the Behavioral Risk Factor Surveillance System**

Carol A. Gotway Crawford, *Centers for Disease Control*

There is an ongoing need to assess health status at both the state and county levels. Estimates of health status help determine demand for health care and inform health policy decision making. The Behavioral Risk Factor Surveillance System (BRFSS) is one of the world's largest ongoing random digit-dialing telephone survey designed to provide state-level estimates of health status. Starting from 2002, BRFSS introduced The Selected Metropolitan/Micropolitan Area Risk Trends (SMART) to provide direct estimates of certain health indicators for selected metropolitan and micropolitan statistical areas (MMSAs) with 500 or more respondents. As a result, about 300 counties have direct estimates of disease prevalence. However, the cost of obtaining reliable estimates for all 3,141 counties is prohibitory large.

Several approaches for modeling county level estimates from BRFSS have been proposed. We provide an overview of several such approaches and contrast the methods with respect to model assumptions, specification and validity, computational time, and benchmarking. We

present an analytic comparison for a subsample of diabetes prevalence estimates and standard errors to evaluate accuracy as compared to direct 2010 estimates. In addition, we summarize differences in estimates from several approaches where direct estimates are not reliable.

### **Small-Area Estimation of Ineligible Sample Units in a Household Sample Frame**

Josué De La Rosa, *U.S. Census Bureau*

Timothy L. Kennel, *U.S. Census Bureau*

Coverage error comprises undercoverage – missing sample units - and overcoverage. Overcoverage is the inclusion of out of scope units (e.g., an establishment on a household frame), duplicate units, and clustered units (e.g., in a cellphone frame, a cellphone shared between two adults).

To account for the overcoverage within a primary sampling unit (e.g., vacant houses in a county), household surveys increase the sample by the amount of ineligible sampled estimated to be encountered in the field. When planning a survey in a new primary sampling unit (PSU) where the true ineligibility rate is not known, survey designers predict the ineligibility rate.

To predict the amount of ineligible sample units in new PSU, household surveys use a national ineligibility rate as a proxy estimate. Using a national ineligibility has the advantage of small variance, however, the high bias of the estimate is inefficient and can be costly. For example, the vacancy rate in an economically stagnate county may bias the estimated ineligibility rate in an economically developed county. This inefficient inflated sample may lead to wasteful over sampling.

Using demographic covariates to model ineligible sample units in PSUs may improve the efficiency of sample designs. This presentation will demonstrate the results of an experiment using small-area modeling techniques to improve the efficacy of estimating ineligible sample units at various geographic levels in address based sample frames.

### **Simulating Populations to Evaluate Survey Estimates**

William Rene Waldron, *The Nielsen Company*

Brian Wells, *University of Michigan*

Yimei Fan, *University of Maryland*

When conveying the results of any survey, the most natural reporting method is that of the direct survey estimate: (# of positive responses to question X / # of total responses to question X). For surveys with large sample sizes, the direct estimate (DE) enjoys some nice properties and is most sensible – but it's not always optimal. Unfortunately, the stability of the direct estimator breaks down for surveys with small sample sizes (or questions with small universes) relative to X. To remedy this, a multitude of alternatives have been introduced, but which one to use?

Now one may survey the vast small area literature (SAE) and find an estimator that seems well adapted to the situation at hand, but there is also likely to be competing estimators that are just as suggestive. One way to evaluate estimators is to calculate their mean squared error (MSE), but it's difficult to compute without knowing the true population value of X. Instead, we propose the construction of a simulated population to inform the decision of the best (i.e., lowest MSE) estimator.

We proceed to generate a hypothetical population using statistical software. The individual elements may be households, persons, or other entities – whatever is being sampled. Prior survey data and ulterior data sources will be utilized so the pseudo-population values of X mirror the true population. Indeed, the construction may be undertaken so the actual data may just as well have been sampled from the pseudo-population as opposed to the actual population. We will exhibit a simulation study that demonstrated how SAE's stabilized AQH ratings estimates for smaller radio stations. A well-designed simulation can give survey administrators a powerful tool to assess whether an alternative survey estimator will produce more accurate reporting.

## **Assessing the Impact of Recent Increases in NHIS Sample Size on Small Area Estimation of Wireless-Only Prevalence**

Stephen J. Blumberg, *National Center for Health Statistics*

Nadarajasundaram Ganesh, *NORC at the University of Chicago*

The National Health Interview Survey (NHIS) is the most widely cited source for data on the ownership and use of wireless telephones. The current NHIS sample design has a state-level stratification, but the NHIS sample size has been insufficient for direct, reliable annual estimates for most states. As a result, state-level estimates of wireless-only prevalence have been produced using small-area statistical modeling techniques that combine NHIS data with auxiliary data to produce model-based estimates (e.g., Blumberg, Ganesh, Luke, & Gonzales, December 2013). Recently, NCHS started receiving supplemental funding to expand the NHIS sample. In 2011 and 2012, the NHIS samples in about 30 states were augmented by expanding the number of sample addresses in those counties already selected for the survey (rather than increasing the number of counties sampled). Nationally, the sample size of interviewed persons rose from nearly 90,000 persons in 2010 to more than 108,000 persons in 2012. In this talk, we will provide an overview of the modeling technique and examine the differences between modeled and direct estimates of wireless-only prevalence at the state level. We will discuss the methods used to increase the NHIS sample size, and we will explore the impact of this increase on the precision of the direct estimates and their contribution (relative to the auxiliary data) to the model-based estimates. In those less populous states where the sample expansion was concentrated, the increased sample size from 2010 to 2012 resulted in greater “weight” associated with the direct estimates when calculating the modeled estimates and narrower confidence intervals around the modeled estimates. The implications of this and future sample expansions for periodic release of direct state-level estimates from NHIS will also be discussed.

## **Methodological Briefs: Survey Mode**

### **Pre-Notification Strategies for Mixed-Mode Data Collection**

Lisa Holland, *Survey Research Center, University of Michigan*

Mick P. Couper, *Survey Research Center, University of Michigan*

Heather Schroeder, *Survey Research Center, University of Michigan*

With steadily increasing reliance upon mixed-mode designs in data collection, it is important to consider the potential effects of various contact strategies on survey response within the context of those designs. In a world of increased reliance on mobile telephones and devices, it is also important to think about how respondents are being invited to participate in surveys, and what methods are most effective for reaching younger generations.

In a follow-up study of more than 2300 active-duty Soldiers in the U.S. Army who recently redeployed, e-mail invitations were used to invite respondents to participate in either a web survey or a telephone interview. In addition to the e-mail invitation and reminders, pre-notification contacts were also made to elicit cooperation from respondents. Soldiers were randomly assigned to one of four pre-notification conditions: personalized letter, text message, both letter and text, or no pre-notification. We examined the effects of each condition on a Soldier's likelihood to participate in the study, as well as their likelihood to participate in either the web or telephone mode. While the literature suggests that pre-notification letters may result in an increase in survey participation (Crawford et al., 2004; Couper, 2008), initial analyses of the present data suggest that the impact on response rates may not justify the cost of implementation. We are currently replicating the study with another 1400 Soldiers. We will report on the results of the pre-notification experiment and the cost implications of implementation.

### **Mailing to Drop Points in a Multi-Mode Survey: Using the NoStat File to Supplement Unit Information**

Katie Dekker, *NORC at the University of Chicago*

Whitney Murphy, *NORC at the University of Chicago*

Drop points are a challenging component of the Delivery Sequence File (DSF) with respect to survey research. Drop points occur when there is a single mail receptacle serving multiple units without unit designation. For in-person studies, synthetic unit numbers are often appended to provide direction for the field interviewer. The lack of unit identifier makes it challenging, however, to pursue mail or phone contact in surveys without a field component. This research will investigate the effect of drop points on two sources of error: coverage and nonresponse. That is, what is the impact on coverage bias if individuals within drop points are excluded from the frame? Separately, does the methodology that we use to include drop points in the frame differentially affect nonresponse and nonresponse bias?

Previous mail studies have taken different approaches to the drop point dilemma. Some studies append a synthetic unit number to the address; others eliminate drop points from mailings altogether. We use data from a multimode (web, mail, and phone) single state general population survey that appended synthetic unit numbers to drop point addresses and included them in the frame. We use the NoStat file, which supplements unit information for a subset of drop points, to enrich our knowledge of drop point unit numbers and attempt a more successful delivery by using knowledge inferred from the supplements to tailor unit numbers in our mailings. We will review the process used in selecting the best unit number to use when mailing to drop points and examine the results of these mailings. We will look at undeliverable rates, web completes, and SAQ returns and compare demographic distributions and key survey estimates of respondents from drop point addresses to those for non-drop point addresses.

### **What's Mode Got to Do With It? Examining Data on School-Age Children Who Speak a Language Other than English at Home**

Angelina N. Kewal Ramani, *American Institutes for Research*

Amber Noel, *American Institutes for Research*

Anlan Zhang, *American Institutes for Research*

The percentage of public school students receiving English Language Learner (ELL) services increased over the past decade (Aud et al., 2013). As the U.S. population becomes more diverse, collecting accurate data on language use and participation in language programs is

increasingly important. However, survey mode may influence whether accurate data on these topics are collected. A non-native English speaker may find it more difficult to complete a mail survey than a telephone interview because there is no person or computer-assisted program available to adjust the survey to a language the respondent speaks (Zukerberg and Han, 2010). This paper will examine the influence of survey mode on estimates of ELL students using data from the National Household Education Study (NHES). These estimates will be compared to administrative data from the U.S. Department of Education's Common Core of Data (CCD). The CCD identifies ELL students in public elementary and secondary schools.

The NHES Parent and Family Involvement in Education Survey (PFI) collects data on education programs and participation, including questions related to language spoken at home by both the child and the parent(s). In addition, parents are asked about their child's participation in English as a Second Language (ESL) programs. The survey was fielded in 2003 and 2007 as a telephone survey, and 2012 as a mail survey. In 2003 and 2007, bilingual interviewers were available to conduct the interviews in English or Spanish. In 2012, English and Spanish screener packages were mailed in targeted areas.

For this analysis, we will compare changes in NHES estimates of the number of school-aged children enrolled in public school and participating in language programs to changes in CCD counts of ELL students between 2003, 2007, and 2012. We will also examine trends in child language use at home and parent-child concordance on language use.

### **Assessing the Impact Device Choice has on Web Survey Data Collection**

Andrew L. Hupp, *University of Michigan*

Heather M. Schroeder, *University of Michigan*

Andrew D. Piskorowski, *University of Michigan*

With the increasing prevalence and ease of accessing the Internet and email through one's mobile device, it is important to assess the impact this method of access has on web survey data collection. 91% of Americans own a mobile device and 52% of them access their email in this manner (Pew, 2013). Therefore, it is becoming more common for potential respondents to view emailed survey requests on their mobile devices and in some cases they proceed directly to the survey. This paper focuses on issues surrounding web survey data collection in a population of primarily younger respondents. We assess the effort expended by both the survey organization to reach respondents and the effort needed by the respondent to complete the survey as impacted by respondent device choice. Survey organization effort is examined through average number and mode of contacts (calls and emails) necessary to reach a respondent and length of time in field. Respondent effort is assessed through number of sessions necessary to complete the interview, breakoffs, and interview length. We evaluate paradata and survey content from the fourth wave of a large study of mental health that investigates the risk and resilience of Soldiers in the US Army.

### **Mode Comparisons, Mode Choice and a Humanized Mode in the European Social Survey**

Marieke Haan, *University of Groningen*

Yfke Ongena, *University of Groningen*

Kees de Glopper, *University of Groningen*

This paper discusses an experiment on the effects of survey mode (CAPI, CATI, text-based Web, video-enhanced Web) and response mode choice on survey response behavior. A



humanized self-administered mode, the video-enhanced Web mode, was developed to investigate the effects of human-like features in Web surveys on data quality. The video-enhanced Web mode contains videos of an interviewer reading questions to the respondent. Mixed-mode data were collected within an additional round of the European Social Survey conducted in 2012 (N=824). Sample members were contacted either face-to-face or by telephone. Half of the sample could choose a mode (the contact mode or Web), the other respondents were randomly assigned to a mode. Half of the Web respondents started with the text-based Web mode and switched to the video-enhanced Web mode half way through the survey, the other half started with the video-enhanced Web mode and then switched to the text-based Web mode. To assess data quality, we analyzed satisficing behavior (non-differentiation, acquiescence, and don't know answers), and studied potential social desirable answering behavior (SDAB) using mixed-effect modeling. Our results show main effects of mode on don't know answering behavior, with most don't know responses in the Web modes. Significant interaction effects of mode/age, and mode/education on don't know answering behavior are found as well. No mode effects are found for non-differentiation, even when comparing Web mode grid questions to other modes. Face-to-face interviewing leads to less SDAB than telephone interviewing. The analyses show the least SDAB in the Web surveys. Significantly more SDAB is found for older respondents and the higher educated. Overall, mode choice and the mode switch in the Web survey hardly affect answering behavior and only minor differences were found between the two Web modes.

### **From Race-to-Face to Mixed-Modes: Lessons from Experiments on the UK Understanding Society Innovation Panel**

Jonathan Burton, *ISER, University of Essex*

Annette Jäckle, *ISER, University of Essex*

Peter Lynn, *ISER, University of Essex*

Understanding Society' is the UK Household Longitudinal Study (UKHLS). Consisting of over 100,000 individuals in 40,000 households at Wave 1, it is one of the largest surveys of its kind. Each year we try to interview all the original sample members, and those living with them, following them as they move within the UK. For the first six waves the data was collected by face-to-face CAPI interviewers. At Wave 7 (January 2015) a mixed-mode design is proposed. Part of the UKHLS is the Innovation Panel (IP). This is a 1,500 household panel sample which is used to run methodological experiments. At IP2 a mixed-mode telephone/face-to-face experimental design was used. At IP5 and IP6 a mixed-mode web/face-to-face experimental design was used. Here we present results from a number of these experiments, giving information about the lessons we've learned - about nonresponse and attrition, the effect of incentives, item nonresponse across modes - which will influence our mixed-mode design and strategy for Wave 7.

### **A Parallel Comparison of AP GfK Poll – Dual Frame RDD vs. KP**

Robert Benford, *GfK*

Curtiss Cobb III, *Facebook*

Jennifer Agiesta, *Associated Press*

Dennis Junius, *Associated Press*

In October 2013, the AP-GfK poll changed both methodology and mode, migrating from dual frame, RDD sample phone surveys to online surveys using KnowledgePanel. KnowledgePanel is an online probability sample frame primarily recruited and empaneled via address-based sampling and prior to that, using random digit dial samples. Prior to the move from one

probabilistic sample to another, GfK conducted parallel surveys using both methods in April 2013. In this paper, we review the similarities and differences in results obtained using the two methodologies via a comparison of the survey demographics, the poll's tracking metrics and timely news content. The research reveals some substantive and significant differences. Past literature has suggested some of these differences are due to mode effects, in particular on attitudes and opinions. But high-profile media polling is sometimes judged on whether its results mirror those of other polls on similar metrics. Differences on these types of metrics for methodological reasons are not necessarily indicative of data quality. There are currently no benchmark measures that can be used for validation of common variables such as job approval ratings. The paper addresses whether the differences found in the parallel surveys are the result of sampling through differential non-coverage and nonresponse or a product of mode effects. The paper also examines whether mode effects found are consistent with past literature. For example, in the April polls, respondents were asked if they support or oppose a legal way for immigrants already in the United States to become U.S. citizens. Results from both samples indicated majority support for a path to citizenship, however, the differences were large enough to potentially lead to differing interpretations of the data. The paper addresses the interpretation of online surveys in the context of other phone polling.

## **Interviewer Observations and Interviewer Ratings**

### **Measurement Effects of Respondent Characteristics on Interviewer-Reported Paradata**

Lindsey Witt-Swanson, *Bureau of Social Research - University of Nebraska-Lincoln*

Interviewers are sometimes asked to provide their observations on aspects of the survey process. These data, referred to as paradata, are collected before an interview to evaluate nonresponse error, but can also be asked after the interview to assess potential measurement error. If these paradata are found to be of high quality, this information could be used to evaluate measurement error among respondents. However, because these data are collected by the interviewer, these paradata could potentially suffer from measurement error also (Blom and Korbmacher 2013; Hill 2002). Measurement error in interviewer observations may arise because of common stereotypes based on respondent characteristics. Research has shown that interviewer characteristics can cause respondents to alter answers to questions related to these characteristics (Davis and Silver 2003; Davis 1997). Research has also found that respondents can bring stereotypes into the interview (Davis and Silver 2003; Marx et al. 1999). If interviewers do the same, paradata quality is questionable.

Data from the 2008 American National Election Study were used to determine if respondent characteristics influenced paradata collected to evaluate measurement error. Five paradata measures were examined around which stereotypes exist by gender, race and/or socioeconomic status, including political knowledge, suspicion of the survey, interest in the interview, respondent intelligence, and sincerity of answers. Respondent gender, race and income, interviewer gender and race, and interaction terms between respondent and interviewer gender and race were used to predict each paradata measure of interest. The results indicate that respondent characteristics influence these interviewer-reported paradata in some cases (e.g., income for political knowledge, interest, and intelligence) but not others (e.g., race for political knowledge and suspicion). Also, paradata were not affected by whether the interviewer and respondent shared a characteristic. Lastly, interviewer characteristics also influence their reports on these paradata. The paper ends with implications for future research.

## **Interviewer Accuracy in Recording Respondent Gender in Telephone Surveys**

Frauke Kreuter, *University of Maryland*

Stephanie M. Kafka, *Gallup, Inc.*

Jenny Marlar, *Gallup, Inc.*

Previous research has demonstrated that interviewer-recorded gender in survey research can be inaccurate (McCullouch and Kreuter, AAPOR 2013). Gender is often used by survey researchers as a key variable in weighting and analysis. Thus, it is critically important that respondent gender is accurately measured.

This experiment quantifies the frequency with which respondent gender is misrecorded by interviewers who code but do not ask gender in telephone surveys. Respondents were assigned to one of three treatment groups at random. In treatment one, interviewers coded respondents' gender early in the interview, directly following the introduction, and then asked the respondent explicitly for their gender. In treatment two, interviewers coded respondent gender at the beginning of the interview and asked the respondent their gender at the end of the interview as a part of the set of demographic questions. In treatment three, interviewers coded the respondents' gender at the end of the survey and then asked respondents directly for their gender. In all instances, interviewers were prompted by the CATI system to record the respondents' gender before explicitly asking if the respondent is male or female. This allows comparison of interviewers' perceptions of respondent gender to actual respondent reports.

The results of this experiment provide survey researchers important information about how frequently respondent gender is inaccurately recorded by interviewers. The results also demonstrate how length of exposure to the respondent affects interviewer accuracy in recording respondent gender. Qualitative interviewer feedback also provides insights into the interviewer experience of collecting gender information from respondents.

## **Evaluating the Association of Initial Expressions of Reluctance with Panel Survey Behavior**

Brett E. McBride, *U.S. Bureau of Labor Statistics*

Lucilla Tan, *U.S. Bureau of Labor Statistics*

Brandon Kopp, *U.S. Bureau of Labor Statistics*

Research has shown that sample units' initial expressions of reluctance at the time of the survey request can be used to correct for later nonresponse through post-survey nonresponse adjustment (Bates, Dahlhamer, and Singer 2008). Further, this reluctance can be classified into distinct themes which help identify whether certain sample units' concerns have unique meaning for understanding their subsequent survey behaviors.

In this exploratory study for the Consumer Expenditure Interview Survey (CE), we examined a set of "doorstep concerns" that sample units could express during recruitment. These 21 items, assigned by interviewers in the Contact History Instrument, were summarized into an overall reluctance score using factor loadings from a principal components analysis to weight each sample unit's reluctance reports across all five survey waves. In addition to classifying sample units by reluctance themes, we studied the association of sample units' rankings along a "doorstep concerns continuum" with their behavior during the survey panel. Three associated survey measures of interest were (1) a sampled person's subsequent response pattern in the survey panel, (2) respondent reporting quality during the interview, and (3) respondent

perceptions of survey burden after the interview. Using bivariate analyses, we found that reluctance ranking groups were effective in differentiating sample units on these three survey measures. Ranking groups were effective in predicting panel response patterns, the presence of preferred survey behaviors (e.g., consulting recall aids), selected data quality measures, and perceptions of survey burden. We also found that these ranking groups were differentiable by the reluctance themes. This presentation will highlight these findings and considerations for the use of these measures in the CE.

## **We Can Hear Where You Are: Evaluating Interviewer Intuitions About Whether Cell Phone Respondents' Are At Home Or Away From Home**

Becky Reimer, *NORC at the University of Chicago*

Daniel Malato, *NORC at the University of Chicago*

Christopher Ward, *NORC at the University of Chicago*

Jenny Kelly, *NORC at the University of Chicago*

Trevor Tompson, *NORC at the University of Chicago*

Jennifer Benz, *NORC at the University of Chicago*

As the proportion of U.S. households that are reachable only or mainly by cell telephone continues to rise over time, it is increasingly important for telephone surveys to include cell telephone samples. Cell phone interviewing presents unique challenges to researchers. Recent research (Ward, et al., forthcoming; Lavrakas, et al., 2010) suggests that substantial numbers (more than 25 percent) of cell phone respondents who complete interviews do so while in locations away from home, and that these respondents may be less willing to provide information on certain types of questions. It stands to reason that, in addition to responding differently during the interview, respondents who are away from home may also behave differently during the screening process and have different cooperation rates than those who are at home. Investigating this research question is potentially challenging given that many cell phone call attempts end quickly and asking every respondent for his/her location on every call is not feasible. Using data collected by the AP-NORC Center, this research examines whether interviewers can use respondent and situational cues to estimate respondent location. For all cell phone call attempts where human contact occurred, interviewers were trained to assess whether the respondent was at home or away from home. On longer calls, interviewers also asked respondents directly about their location, which allowed for the comparison of interviewer perceptions and respondent reports about location. We examine the accuracy of interviewer estimates as well as differences in screener completion and interview completion for respondents believed to be at home versus away from home on the first call where contact was made. We discuss implications for callback strategies and refusal aversion techniques based on these differences.

## **Interviewer Perception of Survey Data Quality**

Yichen Wang, *NERA Economic Consulting*

Brady West, *University of Michigan*

Mingnan Liu, *University of Michigan*

Survey data quality is one of the most important concepts in the total survey error framework and it has received continuous attentions from survey researchers. A series of data quality indicators have been developed to measure data quality and different interviewing conditions have been shown to foster or dampen poor data quality. However, interviewer perception of survey data quality, albeit commonly included in the survey instrument, has received relatively

little attention. Thus the purpose of this study is to determine whether interviewer perception is a valid indicator of data quality. Utilizing data from the Health and Retirement Study (HRS) 2010 Core Survey, this study fitted multi-level multinomial logistic regression models to predict interviewer perceived data quality using a few widely used data quality indicators and interviewing conditions. The models also include both interviewer and respondent characteristics, which allows us to examine whether certain interviewers are good at determining some aspects of the data quality while others are using some other indicators to come up with judgments. Also, we examine whether interviewers have their own patterns in perceiving data quality.

## **Evaluating and Improving Quality of Measurement**

### **Comparability of Common Items across Survey and Administrative Datasets**

Paul Scanlon, *National Center for Health Statistics*

Does a “no” to a question in one survey mean the same thing as a “no” to a similar question in a different survey? As more and more survey and administrative records-based datasets are linked to conduct analysis and capture the power of “big data,” small differences in meanings and interpretations between these datasets are becoming more important. While cognitive interviews are used to pretest an increasing number of both public and private surveys and forms, oftentimes the focus of the cognitive interviews is on new and potentially confusing items and their answer categories. This is problematic as items gathering information on more common constructs are more likely to be used as linking and comparison variables between datasets. Here, we present the findings of a cross-dataset examination of the results of cognitive interviews of common questions and answer categories. We use cognitive interviewing data from projects conducted by the National Center for Health Statistics (NCHS), the Bureau of Labor Statistics (BLS), and the Census Bureau—including the National Survey of Long Term Care Providers, the WHO Modal Disability Survey, the Federal Statistical System Public Opinion Monitoring Survey, the Hate Crimes supplement to the Unified Crime Report, and the Decennial Census/American Community Survey Planning Program—to show that the interpretations of similar questions and answer categories across different survey environments can be widely divergent. Our analysis indicates that one of the largest areas of cross-dataset divergence is the functioning of the common yes/no/don’t know answer categories. Respondent characteristics, such as occupation, educational attainment, as well as survey-specific framing effects, appear to contribute to respondents’ interpretations. Some cross-dataset patterns of interpretations have emerged from this analysis, which should be helpful for survey methodologist and data scientists when deciding whether and how to combine and link data from various sources.

### **Characterization and Analysis of Duplicate Responses in a Physician Survey**

Julianne Payne, *Social & Scientific Systems, Inc.*

Julie Linville, *Social & Scientific Systems, Inc.*

Paul C. Beatty, *National Center for Health Statistics*

Nicholas A. Holt, *Social & Scientific Systems, Inc.*

Researchers sometimes accept survey responses from proxy participants, or individuals presumed to be informed regarding target participants’ knowledge and/or opinions. Proxy responses may be particularly useful in studies of hard-to-reach populations because proxy respondents may be easier to reach than targets. However, the relative quality and quantity of information provided by proxies is not always known. A recent survey, in which responses were

sometimes received from both the target respondent and a proxy, provided the opportunity to compare completeness of responses as well as the extent that proxy responses agreed with those of the target respondent. The National Electronic Health Records Survey (NEHRS), a nationally-representative survey of electronic health record usage among physicians, conducted primarily by mail by the National Center for Health Statistics, provided the data for analysis. From 2011 to 2013, the NEHRS received a growing number of “duplicate” questionnaires from both target and proxy participants (and occasionally, multiple responses from the target respondents as well). We present analyses exploring variation in physicians’ and staff members’ responses, using within-participant duplicates (e.g., physician/physician) as a point-of-reference. We conclude with an evaluation of whether the current NEHRS policy of keeping the most complete questionnaire when duplicates are received is the optimal approach for maximizing data quality.

### **Making Use of Proxy Reports in a Telephone Survey**

Rebecca Medway, *American Institutes for Research*

Celeste Stone, *American Institutes for Research*

Authorizing a proxy respondent to report on behalf of the target sample member is an attractive option for increasing response rates when survey practitioners are unable to make contact with the intended respondent. However, this response rate increase may come at the cost of greater measurement error if proxy respondents are not sufficiently knowledgeable of the sample members about whom they are reporting.

We will report on the use of proxy respondents in a recent telephone survey aimed at determining the feasibility of collecting national estimates of the prevalence of educational and work-related credentials among the U.S adult population. Proxy responses were obtained in two ways. First, if a target sample member was not reached after a certain number of call attempts, another household member was asked to provide a proxy response for the target sample member. Second, the survey included an experiment in which a portion of respondents were randomly selected to receive a request to provide a proxy response for a second household member right after they had reported about themselves.

We will discuss the effect that utilizing proxy respondents had on the response rate. We also will review the characteristics of (1) the individuals who required a proxy respondent and (2) those who agreed to provide a proxy response – as well as whether or not permitting proxy responses improved the representativeness of the final sample. Finally, we will compare the quality of the proxy reports to that of the self-reports by comparing the extent of “don’t know” responses and the relative accuracy of key survey responses as compared to record data.

### **Reconciling Public Participation Rate Differences in Census Bureau vs. Academic/Commercial Survey Estimates**

John Robinson, *University of Maryland*

Tim Triplett, *The Urban Institute*

Relatively small (less than 0.1 percentage point) differences in Census Bureau figures can result in dramatic changes in public policy and planning. When the Survey of Public Participation in the Arts (SPPA) was established in 1982, it was conducted by the Census Bureau to provide the most accurate possible benchmarks of Americans’ participation in the arts in the context of other free-time activities. The Census arts participation estimates in SPPA

1982 were quite consistent with those in subsequent SPPA surveys conducted by Census in 1985 and 1987. However, when the 1997 SPPA was conducted by a commercial survey firm (Westat), rather than Census, the arts participation estimates were consistently and often notably higher than in the previous three surveys conducted by Census. This result obtained consistently in Census SPPA of 2112 (as it had been SPPA surveys conducted by Census in 2002 and 2008). For each arts activity, the Census figures have been up to 13 percentage points lower than those from Westat and other respected independent surveys, like Gallup and the General Social Survey. This article explores several explanations for these lower Census Bureau participation rates, both from internal factors (response rates, activity intercorrelations, survey introduction, etc.) and external estimates – in relation to other academic/commercial surveys conducted by the General Social Survey (GSS) and the University of Maryland. Higher participation rates were also found for a variety of other (non-arts) free-time activities in the SPPA as well.

### **Decreasing Satisficing in Web-Surveys - Evidence from an Awareness Control Experiment**

Sebastian Lundmark, *University of Gothenburg*

Stefan Dahlberg, *University of Gothenburg*

Johan Martinsson, *University of Gothenburg*

Satisficing is a well-known data quality problem throughout all survey research. The introduction of web surveys has generated great opportunities to control and collect data on how respondents answer question, how much time they devote to each question, and even tracking eye movement measuring what words the respondent actually reads. However, in web surveys, respondents might be more easily distracted by other things happening through the same media outlet as where the respondent is supposed to answer the survey (e.g. social media, e-mail). This might increase the likelihood of satisficing and speeding through the survey. One way to decrease this likelihood is to introduce various “speed controls” or pauses in the survey.

In this paper, we attempt to reduce satisficing behavior by introducing an “awareness control” question that has the same outline as other attitude questions, but instead of asking for a response on an attitude, requests the respondent to answer a specific category on a specified scale. Using a 2x2 full factorial experimental design, our study compares the data quality between four groups, one not receiving the awareness control, one receiving it at the end, one receiving it at the beginning, and one receiving it both at the beginning and the end. Preliminary results show a significant increase in data quality in the groups receiving the awareness control early in the survey.

## **Exploring Public Opinion on Societal Issues**

### **The Power of Disinformation: Modeling the Strength and Stability of Institutional Trust in Scientists as the Key Determinant of the Public’s Acceptance of Global Climate Change**

Marc D. Weiner, *Bloustein Center for Survey Research, Rutgers University*

To better understand the attitude formation mechanics behind climate change denial, I develop and test a parsimonious prediction model of the public's acceptance of the fact of hazardous anthropogenic global climate change. I then assess and confirm the centrality to that model of institutional trust, particularly that scientists understand the dynamics of global warming. Finally,

I test and nullify the hypothesis that high levels of recall of extreme weather events generally moderate the impact of trust. These modeling exercises reveal that it is possible to extract a single latent measure of the propensity to accept the reality of global climate change from a set of survey questions that probe agreement/disagreement responses to whether global climate change is (a) occurring, (b) anthropogenic, and (c) a near and/or distant hazardous risk. Moreover, they show it is possible to parsimoniously model the American public's global climate change acceptance propensity using only institutional trust and party identification measures as primary predictors. This lays bare the notion that it is not science, but politics and disinformation that primarily drive the American public's willingness to accept the reality of global climate change. On this issue, recollection of personal experience is trumped, here even for a sample directly affected by (and surveyed only 3.5 to 4.5 months after) Hurricane Sandy, which the New Jersey state climatologist designated the worst storm to strike New Jersey in modern history. While the institutional trust finding is the most powerful, subsample analyses suggest that recall of extreme weather events may, under some circumstances, serve as a useful flashpoint for informing the public conversation on global climate change, and that such communication efforts would do well to highlight the link between global warming and the increasing intensity and frequency of extreme weather events.

### **Parents' Attitudes on the Quality of Education in the United States**

Nicole E. Willcoxon, *AP-NORC Center for Public Affairs Research*

Matthew Courser, *AP-NORC Center for Public Affairs Research*

Emily Alvarez, *AP-NORC Center for Public Affairs Research*

Jennifer Benz, *AP-NORC Center for Public Affairs Research*

Trevor Tompson, *AP-NORC Center for Public Affairs Research*

Improving the strength of public education—particularly early childhood education—is a key feature of President Obama's plan to boost the middle class and grow the economy. Because children who attend preschool are more likely to do well in school and have successful careers than those who do not, the President has proposed creating federal-state partnerships to increase access to high-quality preschool and full-day kindergarten. However, there is great variation in the availability of pre-kindergarten at the state-level—from five states offering universal public pre-kindergarten—to ten other states with no public programs. A key question is whether the availability of state-funded pre-kindergarten programs and per-pupil spending on these programs affect perceptions about education quality and support for publicly-funded universal preschool.

The AP-NORC Center, with funding from the Joyce Foundation, conducted a representative national survey with 1,025 adult parents of children enrolled in grades K-12 during the 2012-2013 school year to measure their opinions about school and teacher quality, the value of preschool programs, and public funding to make preschool available to all 4-year olds. While the survey found that large majorities believe preschool programs improve student performance, and favor using public funds for universal preschool, rates of preschool attendance vary sizably by socioeconomic status. By merging state-level contextual variables with the survey data, this paper addresses whether state preschool and kindergarten programs impact parents' perceptions about education quality and support for a federal preschool program. How do the opinions of parents in states with no preschool access compare to those with high levels of access, and are opinions shaped by the amount of resources allocated to these programs? This study utilizes a unique national survey of parents to understand how state-level policies can affect views and policy preferences concerning America's educational system.



## **How Americans Are Planning, or Not Planning, for Aging and Long-Term Care**

Trevor Tompson, *NORC at the University of Chicago*

Jennifer Benz, *NORC at the University of Chicago*

Rebecca Reimer, *NORC at the University of Chicago*

U.S. Health and Human Services projections estimate that 70 percent of Americans who reach the age of 65 will need some form of long-term care in their lives for an average of three years. With the aging population come important social and public policy questions about preparing for and providing quality long-term care. With a focus on generating new and actionable data about this population to inform the national dialogue surrounding long-term care issues, the AP-NORC Center, funded by The SCAN Foundation, conducted 1,019 interviews with a nationally representative sample of adults who are at least 40 years old. The findings show that although Americans 40 years or older report several concerns about aging and losing their independence, they are not taking actions to plan for their own long-term care needs. There are widespread misperceptions among those 40 or over surrounding the costs of long-term care services, with significant proportions of the population underestimating the costs of nursing home care and overestimating the role of Medicare in paying for that care. Americans 40 or over count on their families to be there for them as they age, but those who are currently receiving long-term care or who have received it in the past are less likely to believe they can rely on their family in a time of need. The survey reveals majority support among Americans 40 or older for some public policy options for financing long-term care. This includes bipartisan support for tax incentives to encourage individual saving for long-term care expenses.

## **To Frack or Not to Frack?: An Examination of Public Opinion of Hydraulic Fracturing in New York and Pennsylvania**

Christopher P. Borick, *Muhlenberg College Institute of Public Opinion*

Barry G. Rabe, *University of Michigan*

As hydraulic fracturing, or fracking as it is commonly known, has become an increasingly common practice through out the United States, state governments have established a wide array of policies related to the regulation and taxation of natural gas extraction. While almost all states that sit above one of the nation's shale plays have enacted some form of legislation related to fracking, the scope and intensity of the policies varies greatly from state to state. Perhaps nowhere is the difference in policy approaches to fracking more pronounced than between New York and Pennsylvania. These neighboring states that sit above the gas-rich Marcellus shale deposits have taken dramatically different approaches to the regulation of fracking within their borders. While New York has maintained a five year moratorium on fracking its neighbor to the south has enacted one of the more industry friendly policies with the passage of Act 13 in early 2012. This stark difference in policy approaches by adjacent states over the same shale play raises the question why such variation exists? In this study we examine the role that public opinion has played in the varied fracking policy adoption in New York and Pennsylvania. Through the use of simultaneous public opinion surveys in both states we measure public knowledge, acceptance and issue saliency and then compare these measures with the actual policies adopted in Harrisburg and Albany. We also examine the views of New Yorkers and Pennsylvanians regarding their knowledge and perceptions of fracking policies in the neighboring states and how these views may shape opinions to fracking within their own state.

## **A Collaborative Assessment of Voter Attitudes and Intensity in New Jersey, New York and Virginia**

Don Levy, *Siena Research Institute*

Harry Wilson, *Roanoke College*

Dave Redlawsk, *Eagleton Poll, Rutgers*

Ashley Koning, *Rutgers University*

Meghann Crawford, *Siena Research Institute*

Working collaboratively, three academic polls – Rutgers-Eagleton, Roanoke and Siena – designed and simultaneously fielded the identical survey in New Jersey, Virginia and New York. Probing voter attitudes towards domestic policies, social issues, fiscal initiatives, foreign policy and quality of life, this panel presents a unique comparison of three states by pollsters that regularly measure public opinion in those areas.

The panel presentation includes an in-depth discussion of 1) building a collaborative effort of this magnitude with the requisite resolution of survey design, sampling, interviewer training, data coding and weighting, and analysis; 2) the questionnaire design using 24 issue indicators across domestic and foreign issues that allows for multiple forms of analysis including logistic regression, factor and cluster analysis as well as determinations of respondent intensity and its effect; 3) substantive findings across multiple issues including same-sex marriage, reproductive rights, social inequality, immigration, healthcare, public education, fiscal policy, privacy and surveillance, and foreign policy; 4) an analysis of attitude direction and intensity both across and within the three states and by key demographics such as party, gender, income and age.

This panel provides insight into voters of three key states that will impact our democracy as incubators of potential presidential candidates, as well as being home to varied and diverse populations that represent broad samples of America. At the same time, this collaboration, points out the fruit, as well as the potential pitfalls of multi-center collaboration. Lastly, this research provides an important multi-state assessment of how measurement affects the expression, dissemination, and interpretation of public opinion, revealing that measurement plays a large role in how policymakers interpret the public, how the public interprets itself, and how we can arrive at a more accurate form of public opinion – and thus a more democratic contribution – when we consider the role of intensity.

## **Responsive Design Paradata Improve Quality**

### **Practical Implementation of Adaptive Design in CATI Surveys – Can an Adaptive Designs Really be ‘Better, Faster and Cheaper’?**

James J. Dayton, *ICF International*

John Boyle, *ICF International*

Naomi Freedner, *ICF International*

Improving survey quality while collecting survey data more efficiently and at lower cost is an implicit goal among all survey researchers and even more critical in the current federal budgetary environment. Adaptive survey design has shown promise in achieving these goals by mid-course corrections of field protocols based on monitoring of key indicators to make evidence based decisions which target specific subsets of the sample. Adaptive design is particularly useful with large sample conducted over time, which permit monitoring, intervention and evaluation of protocol changes.

The Behavioral Risk Factor Surveillance System (BRFSS) is the world's largest, on-going telephone health survey system. Conducted in partnership with CDC, the 50 states, the District of Columbia and four US territories, the BRFSS completes landline and cell phone interviews with over 500,000 households every year. The survey provides state-level estimates for household and individual demographics, as well as a variety of health risk behaviors. The size and scope of the project will allow researchers to study risk behaviors and health program impacts across multiple subgroups.

ICF is identifying study design features that significantly impact cost and response error and we are tracking these indicators during the initial phases of data collection. These indicators include contact rate, cooperation rate, production rate, item nonresponse, sample proportions for key subgroups of interest, estimates of key variables, and paradata. We are looking to optimize call attempts by subgroups based on achieving stable estimates for key variables rather than simply enforcing a minimum number of attempts for all sample records.

The presentation will explore numerous adaptive design features, a pre-identified set of decision rules, as well as the alterations to our data collection approach shown to positively impact data quality and efficiency.

### **Practical Application of Responsive Design in Household Travel Surveys**

Reyna J. Peña, *Abt SRBI*

Jason Minser, *Abt SRBI*

Mindy Rhindress, *Abt SRBI*

Laurie Wargelin, *Abt SRBI*

Randall ZuWallack, *Abt SRBI*

Responsive design in survey research has received considerable attention over the past decade. Responsive design has been positioned as a solution to the problems of increasing survey costs, declining response rates, and coverage error in survey results. As survey researchers continue to face these problems, tested and proven solutions will remain in demand.

Past research has focused on one or more adjustments: in sampling design, incentives, or method of data capture. Rarely have these components been addressed and measured in terms of their interrelated and comprehensive effect on study outcomes. While responsive design in household surveys is not new, the use of a responsive design framework in two-phase household travel surveys is a novel approach and helps in advancing industry understanding of this application. In this presentation, the authors consider practical applications of responsive design components and their interactions, both in Phase 1 and Phase 2 of a regional household travel survey.

The authors examine three recent examples of responsive survey design strategies applied to household travel survey research: 1) sampling considerations; 2) use of incentives; and 3) method of data capture. Benefits and drawbacks to these strategies are evaluated, including increased capture rates for urban households through corrective sampling actions, the marginal effectiveness of tailored incentive structure, and capture rate fluctuations by response method as applied in a two-phase household travel survey. The authors conclude with a discussion of implications and additional research questions.

## **The Use of Paradata (In Time Use Surveys) to Better Evaluate Data Quality**

Ana Lucia Cordova Cazar, *Gallup, Inc., University of Nebraska-Lincoln*

Robert F. Belli, *Gallup, Inc., University of Nebraska-Lincoln*

To improve data quality, researchers evaluate, control, and compensate for errors that arise from different sources. Additionally, error prevention and quality improvement can only be achieved by carefully considering the processes that generate the data (Biemer and Lyberg, 2003). Researchers are increasingly using paradata as an additional tool that assesses the impact of data generation on data quality (Kreuter, 2013). Although informative, paradata usually involve very large and complex datasets, as the units of analysis (keystrokes, mouse-clicks) are nested within higher levels (respondents, interviewers). Such clustering needs to be correctly considered for paradata to validly inform data quality assessment.

In this paper, I aim to improve data quality assessment by using substantive answers and paradata from the 2010 American Time Use Survey (ATUS). Specifically, through the use of multilevel latent trait analysis, a data quality index that will simultaneously account for the various sources of error will be built. The relationship between data quality and interview difficulty will then be explored. The data quality index includes substantive variables about different types of ATUS errors (e.g. refusals and failures to remember an activity), and paradata variables (e.g. response latencies, break-offs). Interview difficulty is measured through paradata (number of entries per activity, number and type of prompts, answer changes). Preliminary analyses do indicate a complex pattern of associations between interview difficulty and errors in the actual ATUS that needs to be disentangled. For instance, two paradata prompts (one related to confirming the accuracy of an activity length, and the other with whom else the activity occurred) are related to two ATUS errors (reporting an activity that couldn't be classified because of insufficient detail, and a failure to report eating). Only one of these prompts, however (activity length), is significantly predictive of another ATUS error, that of failing to remember an activity.

## **The Utility of Alternative Commercial Data Sources for Survey Operations and Estimation: Evidence from the National Survey of Family Growth**

Frost Hubbard, *Institute for Social Research*

Brady T. West, *Institute for Social Research*

James Wagner, *Institute for Social Research*

Haley Gu, *University of Michigan*

Continuing recent declines in the response rates of cross-sectional surveys combined with increases in data collection costs have led to an increase in the use of responsive survey design strategies and consideration of alternative auxiliary data sources for improvement of survey efficiency and post-survey nonresponse adjustments. Among the myriad sources of auxiliary data that might be appended to a sampling frame to support these efforts, commercial databases maintained by consumer marketing or credit scoring organizations have recently received some research attention, due to the rich amount of information that they make available (in some cases at the household address level) for survey researchers. Unfortunately, the existing literature raises concerns about the quality of the information in these databases, and few studies to date have considered the ability of these auxiliary data sources to effectively predict key survey outcomes, including household eligibility for a given survey, calling effort, propensity to respond, and key survey measures of interest. Such studies are important to justify the costs of purchasing these commercial data for survey operations or survey production purposes. With this study, we examine the predictive ability of two alternative commercial

databases offering different sets of auxiliary variables to survey researchers. Analyzing survey outcomes from the U.S. National Survey of Family Growth (NSFG), we find that variables shared in common between the two data sources do have the potential to improve prediction of survey eligibility, but that neither data source substantially improves models of response propensity or key NSFG variables that already include selected NSFG paradata. These results suggest that commercial data can be useful for selected survey operations, but not necessarily for post-survey nonresponse adjustment. Suggestions for practice and directions for future research are provided in conclusion.

## **Applications of Social Media to Survey Research**

### **Effectiveness of Using Facebook Ad Campaign in Collecting Survey Responses**

Shu Duan, *The Nielsen Company*

Jeff Scagnalli, *The Nielsen Company*

Michael W. Link, *The Nielsen Company*

Leah M. Christian, *The Nielsen Company*

Researchers are increasingly turning to opt-in panels, online social networking sites and other non-probability methods for sampling and recruiting research participants. Although these methods lack the ability to provide precise estimates within a certain amount of confidence, they can be especially useful for recruiting participants and collecting responses quickly, at a low cost and for targeting respondents with particular characteristics. Facebook, one of the most dominant social networking sites, has been used to recruit participants for cognitive interviews, focus groups and for experimental research panels (Head, B. 2012 & Autoun, C. et al. 2013). But very little research addresses how effective Facebook is for recruiting participants for completing a one-time survey and what issues need to be considered when designing a recruitment strategy leveraging Facebook's Ad Campaigns. One key advantage of this approach is that it allows survey researchers to target their sampling and recruitment strategies to users with particular characteristics.

Nielsen conducted an experimental test in October 2013 using Facebook to recruit participants to respond to a consumer based survey. The study will examine whether different the placement of the recruitment ads (in the news feed versus right column) impacted participation. The results will be used to assess 1) characteristics of the respondents from Facebook, 2) survey response propensities of Facebook users in general and in different demographic groups, 3) the impacts of using different campaign strategies on cooperation and 4) the time and cost effectiveness of the using Facebook to recruit survey participants.

### **Harnessing Social Media in Survey Research**

Amanda L. Skaff, *Mathematica Policy Research*

Lisa K. Schwartz, *Mathematica Policy Research*

Dan O'Connor, *Mathematica Policy Research*

Social media websites provide important opportunities for respondent outreach in the changing landscape of survey research. With over one billion users, Facebook is a highly popular website that warrants consideration as a resource for improving sample member contact. Facebook is free, easily accessible, and widely-used across demographic groups. Furthermore, Facebook profiles remain constant, even for individuals who are highly mobile and, accordingly, potentially difficult to locate. In an attempt to optimize response rates and retention over time, the

Evaluation of the YouthBuild Program research team is incorporating Facebook into its data collection techniques. In this paper, we assess the ways in which Facebook can augment traditional outreach strategies such as mailings and phone calls. We describe our approach to incorporating social media into our data collection strategy, and detail the benefits of using Facebook to locate and contact this study's sample members, who are generally young, low-income, and highly mobile. We describe the characteristics of sample members who are using Facebook and discuss the kind of information available through Facebook and its value in survey research. Lastly, we analyze differences in demographic characteristics, response rates, and mode of completion between sample members who are our Facebook "friends" and those who are not. Though our social media efforts are ongoing, our research to date suggests that Facebook can be a critical tool for establishing connections with difficult to reach sample members. Establishing Facebook friendship can provide otherwise inaccessible contact information to the study team, helping to boost response rates in current and future rounds of data collection.

### **#DesperatelySeekingRespondents: Using Social Media to Find and Engage Respondents in Social Science Research**

Kelly N. Foster, *East Tennessee State University*

Stephanie Elliott, *East Tennessee State University*

The extent to which social media platforms may be used for survey research recruitment and engagement is an ongoing area of research. This project seeks to explore the feasibility of using both Facebook and Twitter as a recruitment tool to get participants to complete a web-based survey about the use of technology in higher education classrooms. Additionally, the survey asks how participants feel about being recruited for studies via Facebook and Twitter, which, even though they are generally public profiles, often feel private and as such may feel intrusive for potential research participants. For this study, a random sample of students at a regional 4-year university was selected for inclusion into the study. For the most part, these students are part of the Millennial Generation – one of the most technologically engaged group of individuals ever and most of them expect that technology to be part of their college career (McCabe & Meuter, 2011). However, while technology and social media use in the classroom has increased (Taylor, King, & Nelson, 2012) it remains uncertain the extent to which this translates into other realms like research recruitment and engagement. The potential students were then randomized into one of 3 contact conditions – standard mail, email, or social media. All invitation contacts include the same basic information for how to access the survey and their rights as a research participant. Data collection will be completed by March 2014 and analysis will look at response rates by treatment condition, patterns in nonresponse, demographic differences, and respondent's opinions about recruitment strategies. This project seeks to add to the body of literature surrounding new methodologies for survey recruitment by examining not only rates of engagement but also the underlying opinions about the role that researchers should have in engaging potential participants through social media.

### **Connecting: Adapting Survey Methods to Reach Highly Mobile Sample**

Melissa Dugger, *Mathematica Policy Research*

Nan Maxwell, *Mathematica Policy Research*

Temporary housing, frequent moves, gatekeepers, and technological disadvantages (disposable cell phones) are often considered barriers to reaching respondents but are actually the conditions under which many potential survey respondents conduct their daily lives. In 2011, REDF, a San Francisco-based philanthropic organization, contracted with Mathematica to

conduct a study of economically disadvantaged individuals with multiple employment barriers, many of which are also barriers to responding to a survey.

The study included three data collection points: the baseline and exit surveys, which were administered by the organization staff members, and a follow-up interview that was conducted by Mathematica approximately 1 year after the participant started with the organization. Sample members had no further involvement with the organization when their services ended. To facilitate locating, participants were contacted between data collection points. Staff members were hired from the organization as locators, and we also incorporated the use of social media, specifically FaceBook, to locate participants. Using FaceBook, we searched for the participants, using information they provided at the end of the exit survey. After confirming one or more points of information we sent a message to the participant using FaceBook. Weekly monitoring of the date and time of locating attempts, quality of the sample, review of refusals, institutionalized participants or soft appointments, were some of the metrics used to manage the effort and determine when locating attempts needed to be made. Using social media, along with paradata metrics on productivity and locating, we were able to determine the best strategy for locating participants.

In this presentation, we will cover the results of the follow-up data collection and the impact the locating effort had on the final response rate and strategies for re-conceptualizing the approach to finding hard-to-locate populations.

### **A Look at Mobile Device Usage Among College Students**

Jennifer Brooks, *National Survey of Student Engagement (NSSE)*

Shimon Sarraf, *National Survey of Student Engagement (NSSE)*

Cole James, *Beginning College Survey of Student Engagement (BCSSE)*

The widespread adoption of mobile technologies has dramatically impacted the landscape for survey researchers (Guidry, 2012). Each year, increasing numbers of respondents are accessing web surveys on handheld devices and tablets, instead of desktop computers. Some research suggests that there is little or no substantive difference for users accessing web surveys on tablets as compared to desktop machines, but the user experience on a handheld device is recognized to be significantly different than the other modalities (Buskirk & Andrus, 2012). This has caused many researchers to optimize their web surveys for small-screen, mobile technologies. However, little is known about how this may affect survey data or respondent characteristics.

Using results from hundreds of institutions that participated in the 2013 National Survey of Student Engagement (NSSE) the current study details mobile traffic among NSSE respondents, and examines the demographics and engagement results of mobile respondents compared to non-mobile respondents. Are mobile respondents younger, or typically more male than female? Do mobile respondents score differently in NSSE measures than desktop users? Is there any suggestion of selection bias between the two modes?

Approximately 48,000 respondents used a small-screen, mobile device to complete NSSE in 2013. Because NSSE surveys college and university students exclusively, this study provides unique insight into behaviors of a demographic group of particular interest to survey researchers (17-24). The results will lead to a better understanding of mobile respondents and whether they differ from computer-based college-age respondents.

**Saturday, May 17**  
**8:00 a.m. – 9:30 p.m.**  
**Concurrent Session G**

## **The Future of Landline and Cell Phone Telephone Surveys in the U.S.**

### Panelists:

David Dutwin, *Social Science Research Solutions*  
Trent Buskirk, *Marketing Systems Group*  
Courtney Kennedy, *Abt SRBI, Inc.*  
Tom Guterbock, *University of Virginia, CSR*  
Jennifer Kelly, *NORC at the University of Chicago*  
Andy Peytchev, *RTI International*  
Linda Piekarski, *Survey Sampling International, LLC*

The dual-frame (landline/cell phone) RDD sampling landscape continues to change rapidly. Cell phone ownership in the US has reached over 90 percent in just over a decade. In that time, landline non-working rates and refusals rates have been on the rise, leading to less efficiency and higher costs. Yet at the same time there is anecdotal evidence that participation rates on cell phones have been stable or may have even started to increase. Costs for cell phone surveys also have stabilized and may even be decreasing. This appears to have happened despite the fact that calling cell phones remains less productive given the requirement of manual dialing in the US. In addition, the lack of robust auxiliary data makes the cell phone frame less attractive.

So we ask: is the age of landline dialing coming to an end in the US, or will it continue to have a viable role well into the future of survey research? Are productivity rates between landlines and cell phones converging, and if so, (when) will use of the cell frame become the more cost-attractive? And perhaps most importantly, will US researchers be conducting "traditional" telephone surveys of any kind in the future, given current trends?

This panel is comprised of experts who have helped lead research on the trends in US cell phone and landline surveys for the past decade and have endeavored to answer the questions posed above. They include senior personnel from sampling organizations, survey research organizations, and universities with dedicated survey research programs. They include considerable expertise on the operational aspects of dual-frame telephonic research in the US, and those experienced in considering the relative costs of research using both frames, as well as those who have researched the validity of data attained by a single frame—cell phones, only.



# **A Comprehensive Overview of Immigration and Immigrants in the U.S.: Public Opinion, Public Policy, Demographic Change and Immigrant Integration**

## **Message Research on Immigration and Immigration Reform: Introducing a New Message Research Methodology and Reviewing Current Messaging Pro and Anti-Immigration Legislation Reform**

David Mermin, *Lake Research Partners*

This presentation will focus on public opinion and attitudes related to immigration and immigration reform as well as message research around the issue. The messaging research draws on two studies that included national online dial surveys of voters. One study reached a representative sample of all U.S. voters (2012) while the other reached a sample of women voters (2013). Both projects draw on a new methodology that was designed to find new frames and messages that engage the base, persuade the middle, and alienate the opposition. This presentation will also explain the utility of this new methodology and how it can be applied to other research questions.

## **Religion and Attitudes Toward Immigrants and Immigration**

Gregory A. Smith, *Pew Research Center*

Elizabeth Sciupac, *Pew Research Center*

Cary Funk, *Pew Research Center*

Jessica Harar Martinez, *Pew Research Center*

This paper takes up the question of the link between religion and attitudes toward immigrants and immigration. Immigration policy is of perennial concern in American politics. And religion is known to be one of the most powerful forces shaping individual attitudes on matters of public policy and electoral politics. The link between religion and immigration attitudes, however, is often speculated about but not very well understood.

Members of many religious groups may well feel cross-pressured with respect to immigration. Time magazine recently reported, for example, that two influential evangelical Christian groups have “dueling interpretations of the gospels’ guidance” about immigration. And the views and recommendations of religious elites – from Catholic bishops to mainline Protestant and evangelical leaders – are often discussed in news coverage of the immigration issue. But the impact of these efforts by religious elites on the attitudes of members of their respective flocks remains unclear.

In this paper, we analyze the correlation between religion – including religious identification or affiliation, religious behavior, and religious belief – on attitudes toward immigrants and immigration policy. We assess whether some religious groups are more favorable toward immigrants and supportive of immigration reform than others, and explore the connection between overall levels of religious commitment and attitudes on these issues both for the population as a whole and within respective religious traditions (e.g., are evangelical Protestants who attend church regularly more or less favorable toward immigrants as compared with evangelicals who attend church less often?). And using multivariate modeling techniques, we examine whether or not any apparent connection between religion and immigration attitudes persists even when other important variables -- including race/ethnicity, age, gender, education, and geographic region – are taken into account.

## **Comparative Surveys on Support for Democracy and Democratic Attitudes**

### **Public Attitudes Towards Governmental Corruption in Comparative Perspective**

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

Jibum Kim, *Sungkyunkwan University*

Sori Kim, *Sungkyunkwan University*

The International Social Survey Program in 2004 and 2006 asked about the public's perception of 1) corruption among government employees, 2) corruption among elected officials, 3) the need to bribe public officials to get government services, and 4) whether a citizen's treatment by the government depending on personal ties to politicians. This paper looks at differences across countries and across socio-demographic groups within countries and how attitudes towards corruption affect overall evaluations of the political system.

### **The CNEP and the Comparative Study of Democracy**

Richard Gunther, *Ohio State University*

The Comparative National Elections Project (CNEP)—currently consisting of 28 national election surveys conducted in 18 countries—includes batteries of questions that deal with support for and satisfaction with democracy, differing understandings of democracy, and citizen engagement in democratic politics. Accordingly, over the past two decades analyses have been conducted concerning the dimensional structures linking these attitudes, their hypothetical origins and their behavioral consequences. In this paper, we shall see that Democratic Support, Democratic Satisfaction, and Disaffection constitute three conceptually and empirically distinct clusters of attitudes, with different origins and behavioral consequences. We shall also see that in many countries, particularly those that have democratized most recently, majorities of the population conceive of democracy in substantive (and predominantly economic) terms, rather than processes such as elite-recruitment procedures. We shall also see that political parties and elites play crucial roles in consolidating their respective democratic regimes (or not), and in framing popular understandings of democracy.

### **World Value Survey: Support for Democracy in 60 Countries, 2011-2014**

Christian Haerpfer, *University of Aberdeen*

The World Values Survey will conclude in spring 2014 its 6th wave with face-to-face interviews in 60 countries. The paper is using the concept of 'support for democracy' with 6 levels of support from diffuse support for the political community of the nation-state to the specific support for political actors like politicians.

Levels of support for democracy include support for the nation-state, support for democracy and its principles in general, support for the current political regime, support for the incumbent government, support for specific political institutions like parties, pressure groups and civil society as well as support for specific political actors, like Presidents, Prime Ministers and leading politicians in parliament and government. The paper is measuring support for democracy on a global scale and is testing, if there are global trends regarding support for democracy or common trends in specific regions and continents of the world.

## **Generational Differences in Support for Democracy: Evidence from Latin America**

Alejandro Moreno, *Instituto Tecnológico Autónomo de México ITAM*

Marta Lagos, *Latinobarometro Corporation*

In this paper, we examine the extent to which democratic support has grown among younger generations of Latin Americans, who have socialized under more open and competitive political regimes as their elders. We test socialization hypotheses and look at differences in support for and satisfaction with democracy among different generational cohorts using Latinobarometro data, a survey in 18 Latin American countries conducted annually since 1995. We find little evidence of generational gaps, and discuss this in a comparative perspective according to findings from other newly democratic contexts in East European societies.

## **Blurred Lines: Big Data, Ethics, Privacy and Confidentiality**

### **Analyzing Data Sets: The Ethics of Using Big Data**

Kirsten Martin, *George Washington University*

The use of Big Data – the aggregation and analysis of large data sets in order to identify both trends and personally identifiable data – is becoming more popular in commercial and government sectors. However, not all data sources are equal. Data sets should be analyzed to determine not only if the data is accurate and if the knowledge is fit for use, but also if the data stewardship practices of the data source match that of the organization, and if the use of the data source has any harmful consequences. The goal of this paper is to outline guiding principles and questions in analyzing data sets – such as administrative records and third party data – for use in statistical analysis. While the Bureau has been conscientiously analyzing how data is disseminated and used once gathered by the Bureau, the turn to analyzing how data is gathered by alternative sources is less established. This paper is a small step to frame how to assess alternative sources of data sets for use in surveys.

### **Privacy, Big Data and the Public Good: Frameworks for Engagement**

Stefan Bender, *IAB*

“Big data” can be harnessed to serve the public good: scientists can use new forms of data to do research that improves the lives of human beings, federal, state and local governments can use data to improve services and reduce taxpayer costs and public organizations can use information to advocate for public causes.

The goal of this presentation, and the book by the same title, is to answer some of the unanswered questions about using Big Data according to ethical and legal requirements as well as public opinion. How do scientists and government officials seeking to serve the public good proceed without harming individual citizens? What are the rules of engagement? What are the best ways to provide access while protecting confidentiality? Are there reasonable mechanisms to compensate citizens for privacy loss?

The book’s authors paint an intellectual landscape that includes the legal, economic and statistical context necessary to frame the many privacy issues, including the value to the public of data access. The authors also identify core practical approaches that use new technologies to simultaneously maximize the utility of data access while minimizing information risk.

## **Social Media and Voluntary Disclosure of Big Data: Unexpected Consequences of Online Disclosures**

Laura Brandimarte, *Carnegie Mellon University*

Online social media has grown exponentially in the last few years, effectively working as a means of vast data collection and dissemination. What are the consequences of such an amassment of “Big Data,” voluntarily disclosed by social media users, in terms of impression formation and judgment of others?

In this paper, we investigate whether the act of disclosing sensitive personal information to others makes one more understanding or more judgmental towards those who made similar disclosures. Focusing on an online setting and using observational data, in a first study we find that participants who disclosed information about a questionable behavior judged others who did the same more harshly as compared to participants who did not disclose. This result is confirmed in a second, experimental study, where we also tested for the causal effect of disclosure by using instrumental variable estimation.

Our studies show how the use of social media for impression formation, together with evolving social norms about disclosure, which seem to go in the direction of openness and sharing, could have unexpected, negative consequences for the people who disclose. Although social media users are voluntary providers of “Big Data,” they may surprisingly find themselves victims of the very system they are contributing to create. This could cause a change in willingness to disclose in the future.

## **Drones and the Human Element: Public Perceptions and Implications for Survey Research**

Joe Eyerman, *RTI International*

Unmanned Aircraft Systems (UAS) are a relatively mature technology currently used for military and homeland security purposes that are quickly being developed for transition to public safety, first responder, and commercial applications in the United States. Although currently closely regulated by the Federal Aviation Administration (FAA), UAS are expected to be cleared for wider use in the continental United States following the congressional mandate that FAA establish guidelines for use within U.S. national airspace (NAS) by 2015 (H. Rep. No. 112-381).

RTI International began a research program in 2012 dedicated to understanding the social, behavioral, and policy factors associated with UAS technology. As part of this research program RTI conducted two pilot studies to inform our understanding of public perceptions and law enforcement concerns. We have initiated a third project to investigate public health implications of transitions of UAS into the NAS.

This presentation will review the findings from the RTI surveys and other opinion data about the willingness of the US population to accept new technologies that may infringe on privacy and civil liberties.

## Questionnaire Design: Improving Data Quality with Better Questions

### Improving Survey Questions by Finding Alternative Wordings to Low-Frequency Words

Ana Slavec, *University of Ljubljana*  
Vasja Vehovar, *University of Ljubljana*

Developing a good survey questionnaire is a complex task where several decisions need to be made regarding question format and wording. Each question can be worded in different ways and it is important to carefully select the most appropriate words. According to Krosnick and Fabrigar (forthcoming) one of the criteria for good wording is meaning uniformity, that is achieving single meaning for all respondents. Thus, words with many possible interpretations, as well as abstract, ambiguous and emotionally charged words should be avoided. Sometimes, problematic words can be identified by their low frequency in linguistic corpora. In addition, there exist computerized solutions that can help us indicate potential difficulties, for instance QUAID (Graesser et al. 2006). However, these programs do not suggest alternative words, synonyms or hypernyms that are more familiar to the respondent. The researcher has to find them on their own, possibly by using a thesaurus (e.g. Microsoft Word thesaurus) or a lexical database (e.g. WordNet). Moreover, it is not explicitly known if words are interchangeable and what are the effect of replacements. Do they improve question quality? Are the original meanings of the questions preserved? In this paper, we explore the capacity and applicability of thesauruses and lexical databases to find synonyms and hypernyms as replacement words. We perform split-ballot experiments for selected wordings of questions from three different questionnaires, two performed on a student population and one on university employees. On this basis, we evaluate the potential of thesauruses and lexical databases to improve survey questions and future possibilities of integrating them with existing survey software.

### Format Effects in Looping Questions

Stephanie Eckman, *Institute for Employment Research*  
Frauke Kreuter, *University of Maryland, JPSM*

Previous research has demonstrated that the way in which filter questions are asked can affect the responses given: respondents tend to give fewer answers which trigger additional questions when the filters are interleaved with the follow up questions than when the filters are asked all in a group. We extend this research to looped questions in which respondents are asked the same battery of questions about every full-time job they have held, or every degree they have received. Such looping questions are common in surveys which collect biographical histories, but little prior work has explored the best way to ask such questions. Like filter questions, looping questions can be asked in two formats: one which asks first how many full time jobs a person has held, and another which first asks about one job and then asks if the respondent has held another job. We call these two formats “how many” and “go again.” In this paper, we investigate whether the format effect that we find in filter questions also applies to these looping questions. Based on the filter question research, we expected to find reduced reporting in the “go again” format. To investigate the phenomenon, we use data from a recent web survey in German ( $n=1,068$ , AAPOR RR1=10.3%). We do find the expected effect. Exploiting a link between survey responses and administrative data which is available for more than half the sample, we also show that respondents in the “how many” condition give more accurate responses on the number of events, and those in the “go again” condition tend to underreport. However, there may be other reasons to prefer the “go again” format, as it allows respondents

to discuss one event at a time. Our results provide guidance to questionnaire designers, survey practitioners and analysts of survey data.

## **Strategies for Retrieval From Autobiographical Memory**

Erica Yu, *Bureau of Labor Statistics*

Scott Fricker, *Bureau of Labor Statistics*

Questions about past behavior often require respondents to retrieve information from autobiographical memory -- a combination of personal experiences (episodic memory) and general knowledge (semantic). An example would be recalling whether you recently purchased any sweaters and, if so, the price that you paid. The retrieval of autobiographical information is fundamental to the design of survey questions and yet remains poorly understood. We explore these issues in the context of the Consumer Expenditure (CE) Survey. Participants in the present study complete an online task in which they are asked to report on items they have purchased, beginning with an “anchor” item selected at random from a range of CE categories. Upon responding to the first anchor item, participants will be prompted to report the next purchased item that comes to mind. After reporting five items, the participant will be asked to describe the relationship between each consecutive pair of items, both with an open-ended description as well as a choice from a selection of relevant dimensions (e.g., purchased in the same shopping trip, purchased for the same purpose). The text and choices of the relationships between items will then be analyzed to identify common themes relating to retrieval from memory. In particular, whether retrieval is clustered and around what dimensions, as well as which strategies lead to recall output aligned with survey goals. We will discuss whether systematic retrieval strategies exist across individuals, what factors influence the use of different strategies, and how researchers can utilize this knowledge to optimize retrieval of autobiographical information in surveys.

## **Is It Too Much to Ask? The Role of Question Difficulty in Survey Response Accuracy for Measures of Online Behavior**

Kumar Rao, *The Nielsen Company*

Mimi Zhang, *The Nielsen Company*

Tanya Luo, *The Nielsen Company*

While market research capabilities of online panels have never been greater, the challenges facing these panels are just as great. Over the past few years, online panels that recruit members using nonprobability/opt-in based methods have come under increased scrutiny and criticism over data quality concerns such as respondent identity and increased satisficing. These concerns have drawn attention to the heart of the issue, which is: the accuracy or truthfulness of data provided by opt-in panel respondents. This issue is of utmost importance given the recently established link between opt-in panel sample and poor survey data quality (see Yeager et. al. 2011).

In this study, we tackle this issue by assessing the accuracy of multiple survey measures of online behaviors by comparisons to behavioral data collected via respondents' membership with a metered panel. The survey measures were questions around visitations to various websites including a popular social networking website. While research on assessing the accuracy of self-reports has been around for a while (see McMorris & Ambrosino 1973; Miller & Groves 1985) and has generally involved comparing self-reports against administrative/official records, this study is unique since it assessed the accuracy by comparing self-reports against actual behavior. The behavioral data consists of respondents' online activity captured passively through a software meter installed on their primary household computer. Results revealed an

interesting relationship between difficulty of the survey question and accuracy of the response. We found that response accuracy was higher for simple than complex questions on online behaviors. This is particularly the case for most websites, irrespective of their popularity. Furthermore, the propensity of providing an accurate response when controlled for question complexity showed interesting demographic effects of age and cognitive ability. This study concludes that opt-in survey can provide accurate and valid data when done right i.e. with reasonable task expectations.

### **Easy Question, Tricky Answer: Measurement Quality of Education Questions**

Jessica M.E. Herzing, *GESIS Leibniz Institute for the Social Science*

Silke L. Schneider, *GESIS Leibniz Institute for the Social Science*

Education is one of the most frequently used variables in social science research, as it is one of the best predictors for different forms of social inequality. However, the complexity of measuring education adequately has increased in recent decades, due to rising differentiation of educational systems and migration. Consequently, showcards used for education questions have become rather difficult for respondents to answer. Thus, measurement problems in this commonly used variable should be the rule rather than the exception. The question at hand is whether it is possible to replace showcards with open-ended questions, to overcome the problem of incompleteness of fixed choices. Therefore, we analyze data from the GESIS Panel Pilot, which is a German online panel with about 500 respondents. Our assumptions will be tested by the experimental variation of question wording and question type in a split ballot design. We employed four open question variations and one fixed choice question module asking for educational attainment. This provides the opportunity to perform reliability checks of respondents' answers by comparing the education level resulting from the open responses with the answers to fixed choice questions. Furthermore, our exploratory analyses focus on the time taken to answer a question as a function of item level characteristics and respondent characteristics to measure the data quality. To evaluate the measurement quality of education questions in the different splits we address the following questions:

- a) Are different types of questions (fixed choice or four variations of open questions) associated with different item-nonresponse rates?
- b) Are different types of questions (fixed choice or four variations of open questions) associated with shorter or longer response times?
- c) Is data quality, such as test-retest reliability, associated with completion time and length of the question wording?
- d) Is the test-retest reliability associated with different question characteristics (e.g. question specificity, providing examples)?

## **2014 Wave or No Wave - Mid-Term Election Preview**

Panelists:

Mark Blumenthal, *HuffPost Pollster*

Mark DiCamillo, *Field Research*

Celinda Lake, *Lake Research*

Frank M. Newport, *Gallup, Inc.*

Anthony Salvanto, *CBS News*

Floyd Ciruli, *Ciruli Associates*

A panel of leading public and political pollsters will preview the 2014 election and analyze the trends that appear to be driving the contest, both state and national.

Will 2014 be a wave election or is it so choppy no trend is spotted? Could President Obama and the Affordable Care Act give Republicans the national advantage they seek? Does the dislike of Washington and/or Congress provide an advantage to either party? Is the polarization of politics so prevalent in Washington D.C. increasingly visible at the state level with hard left and right moves? How will the arrival of big data and polling aggregators affect the traditional media and candidate pollsters?

## Topics in Survey Mode

### **A “Green” Appeal: Efficacy Evaluation of Assigning Sample Members that Prefer the USPS Mail Mode to the Online Mode in the 2013 Survey of Doctorate Recipients**

Shana M. Brown, *NORC at the University of Chicago*  
Jessica Knoerzer, *NORC at the University of Chicago*  
Lance Selfa, *NORC at the University of Chicago*  
Lynn Milan, *National Science Foundation*

The Survey of Doctorate Recipients (SDR) is a biennial multi-mode panel survey of doctorate recipients in the science, engineering, and health fields from U.S. degree-granting institutions sponsored by the National Science Foundation and National Institutes of Health. The final SDR questionnaire item asks for future survey mode preference, i.e., mail, online, telephone, or no preference. Based on reported “mode preference,” SDR places cooperative sample members, who participated in the prior round of the survey, in one of three start modes for the subsequent round. The default starting mode for cooperative individuals without an explicit mode preference is the response mode from the previous SDR. The 2008 SDR conducted a controlled experiment to determine whether cooperative sample members who implicitly preferred the mail mode could be successfully started in the more cost-effective and timely online mode; ultimately, level of effort and response were not significantly improved for the treatment group. In 2013, the SDR conducted a similar experiment, with cooperative sample members who explicitly or implicitly preferred the mail mode, but this time made a “green” appeal in the first sample member contact for the round. This appeal encouraged the treatment group to participate online rather than by mail for the purpose of being more efficient and eco-friendly. Sample members completing via mail survey in 2010 and had both email and mailing addresses on file were randomly assigned to the online start mode or the mail control group. Analysis will focus on the comparison of treatment and control groups by unit response rate, days to response, and required level of effort. We will also look at the demographics of those for whom this approach was successful. The results of the data will indicate if similar cases can be assigned to the online start mode in future survey rounds.

### **Mode Experiments in Mixed-Mode Surveys: Insight from the Cognitive Economics Study**

Joanne W. Hsu, *Federal Reserve Board*  
Brooke H. McFall, *University of Michigan*

The Cognitive Economics (CogEcon) survey is a longitudinal study administered to a representative sample of older Americans. The study focuses on household finance, economic



decision-making, and financial literacy. The first wave in 2008 included a mode experiment: invitees who indicated in a screening interview that they used the Internet (“web-eligible” invitees) were randomly assigned to complete either a web-based interview or paper questionnaire. The resulting experimental design enables comparisons between web-eligible invitees’ response propensities and item response qualities when invited to a mail versus a web survey. We examine differences in survey response rates, item nonresponse and data quality due to both mode effects (for questions that were presented identically across modes) and web-enabled design innovations (for questions that were designed differently across modes to exploit the superior adaptability of web surveys), particularly for economic questions. Our analyses show that, in contrast to mail mode, the CogEcon web survey: (1) attained higher response rates among web users; (2) displayed higher item nonresponse; (3) gathered more precise values for wealth measures. However, the mode effects appear to be small for questions that were designed to be identical across modes. We conclude that, for web-using populations, web surveys appear to result in more usable data than mail surveys, these data appear to be of high quality, and these improvements offset the higher item nonresponse displayed on the web. In follow-up surveys, we administered web surveys to all respondents with Internet access and retained data quality levels we observed for web respondents in wave 1. Therefore, to strike a desirable balance between cost-effectiveness, representativeness, and bias, we propose a protocol for longitudinal mixed-mode studies consisting of a single-mode screener for all respondents, followed by a first wave with mode randomization and subsequent waves inviting all respondents with web access to a web-based questionnaire.

### **Belt and Suspenders: Evaluating the Efficacy of Sending Initial Contacts via Email Only vs. USPS Letter-Plus-Email to Online Responders in the Survey of Doctorate Recipients**

Ipek Bilgen, *NORC at the University of Chicago*  
Shana M. Brown, *NORC at the University of Chicago*  
Lance Selfa, *NORC at the University of Chicago*  
Lynn Milan, *National Science Foundation*

The Survey of Doctorate Recipients (SDR) is a biennial multi-mode panel survey of doctorate recipients in the science, engineering, and health fields from U.S. degree granting institutions sponsored by the National Science Foundation and the National Institutes of Health. Each round of the survey, sample members are assigned to a start mode of either mail, telephone, or online. Cooperative sample members are assigned based on a question in the survey or their completion mode. In past rounds of the SDR, those assigned to the Web start mode that have a primary USPS address and email on file received their initial contact in a dual mode. The SDR sent a letter with online survey access to the primary USPS address on file with a corresponding email with the same online survey access sent three days later to the primary email address on file. For 2013, we theorized that cooperative panel cases identified to start in the online mode would respond equally well to a single email contact as to the email plus USPS letter combination used in prior rounds. The number of cases starting in the online mode has increased rapidly to about 26,786 cases in the 2010 SDR, so reducing mailouts to this group could realize substantial savings for the project. Cases assigned to the 2013 online start mode with both an email and USPS address on file were randomly assigned to the single initial email contact or the standard USPS letter-plus-email initial contact. Analysis will focus on the comparison of treatment and control groups for differences in response rate, days to response, and “level of effort” measured by the number of USPS mailings sent, telephone contacts administered, and email messages sent. We will also look at the characteristics of those for whom this approach was successful.

## **Response Rate and Participant Differences in a Screening Survey Using Interactive Voice Response (IVR) and Computer Assisted Telephone Interviewing (CATI)**

Julita Milliner-Waddell, *Mathematica Policy Research*

Rebecca DiGiuseppe, *Mathematica Policy Research*

Using telephone interviewers to screen large numbers of potentially study-ineligible individuals is extremely costly. The use of interactive voice response (IVR) provides an alternative for reaching sample members and reducing labor costs. This paper will examine the use of IVR among sample members in nine states who filed for unemployment insurance (UI) benefits during 2010.

The study, being conducted by Mathematica for the U.S. Department of Labor, is evaluating the impact of a subsidy for health benefits under the Consolidated Omnibus Budget Reconciliation Act (COBRA) that was provided by the American Recovery and Reinvestment Act (ARRA) of 2009. The sample was comprised of more than 26,000 UI claimants; however, only those eligible for COBRA benefits were eligible for the study. It was estimated that fewer than 50 percent of the sample would be study eligible.

To screen this large number of individuals, sample members were asked to either call into an IVR to complete a brief screening survey or call the Mathematica survey operations center directly to determine their eligibility for the study. To encourage the use of the less-costly IVR option, sample members were offered an additional \$10 if they were determined eligible through the IVR and completed the full interview.

In this paper, we report on differences in IVR use by several sample member characteristics, including age, race, and region. We also look at differences in time of day completed to determine whether the extended hours offered by an IVR is important for all or some groups, compare full survey completion rates among IVR and non-IVR users, and estimate cost savings provided by use of the IVR as compared to a telephone screening operation only.

## **Effect of Recruitment Mode on Survey Panel Participation, Retention and Response**

Heather E. Driscoll, *ICF International*

Kurt Peters, *ICF International*

Quality longitudinal panel research depends on maintaining panel participation and data quality over time. ICF conducted a longitudinal panel study as part of the National Recreational Boating Survey (NRBS), sponsored by the U.S. Coast Guard to ensure that the public has safe, secure, and enjoyable recreational boating experiences. Specifically, the NRBS Program enables the Coast Guard to better identify safety priorities and coordinate and focus research efforts. The NRBS featured several components, including a "Trip Survey" which sampled a panel of nearly 24,000 boaters over the course of 12 months. The Trip Survey was designed to allow short-term recall of recreational boating trips. The Trip Survey panel has been in place for over a year and was recruited via the dual-frame, dual-mode (Random Digit Dial telephone and mail) "Boat Survey". Potential panelists from the mail frame of state boat registration lists received up to four mail contacts, including a pre-notification and cover letters signed by a US Coast Guard official which provided detailed information about the survey and how the data would be used.

Potential panelists from the RDD frame were read a short introduction, which provided a brief explanation of the survey. Of the 23,881 panelists, 15,864 were recruited from the RDD frame, and 8,017 were recruited from the mail frame.

This research investigates differences in these two panel groups, including initial agreement to be on the panel, participation/retention over the course of the project, and key responses. We also offer lessons learned from recruitment mode to optimize panel retention and data quality.

## **Surveying Them Online: Recent Issues and Current Trends**

### **Exploring Use of Web Surveys for Multi-Generational Studies**

Pamela Giustinelli, *University of Michigan, Survey Research Center (ISR)*

Robert Willis, *University of Michigan, Survey Research Center (ISR)*

Michael Zabek, *University of Michigan, Survey Research Center (ISR)*

The Health and Retirement Study (HRS) is the premier survey of older Americans' work and health transitions. While also covering intra-family relationships, it remains limited to reports by one side. This feature greatly limits research on intergenerational transmission, support-and-exchange, etc. Recent introduction of a web-HRS, however, opens to a unique opportunity for surveying children and parents of current respondents, while circumventing high costs and logistic difficulties of in-person interviews.

We pilot this idea using the American Life Panel (ALP), an online panel of 6,000+ respondents from the adult US population and an ideal test bed:

- (a) Since over half of ALP panelists completed the 2012 web-HRS, the HRS family roster provides us with a sampling frame of their children, parents, and spouse/partner.
- (b) Existing "invitation" and "subscription" protocols (from respondent-driven sampling experiments using panelists as seeds) provide us with a customizable framework for our application.

We design and field

- (1) An Invitation Survey asking a sub-sample of ALP panelists to invite their eligible relatives (i.e., adult children, living parents, and current spouse/partner) from the pre-loaded family roster (a) to join the study. The survey also includes questions about respondents' use of web and communication technologies, in general and to communicate and keep in touch with their family.
- (2) A Referral Survey enabling invitees to join the study.

We follow up

- (1) On inviters with an Invitation Follow-Up Survey, covering (i) invitation status and perceived subscription status for each relative; (ii) subjective probabilistic beliefs over invitation and subscription; (iii) invitees' background characteristics.
- (2) On invitees with a Web and Communication Technology Use Survey with questions analogous to those in the invitation survey.

We derive estimates of invitation and subscription rates by relationship type. We analyze them as a function of relevant characteristics of inviters, invitees, and their relationship, and of design features of the invitation-and-subscription protocol.

## **Who Is Behind That Screen? Solving the Puzzle of Within-Home Computer Sharing Among Household Members**

Tianjue Luo, *The Nielsen Company*

Kumar Rao, *The Nielsen Company*

The number of US households with access to computers at home has continued to grow. According to the 2011 Computer and Internet Use report published by US department of Commerce, 77% of US homes have computers in their home, compared to 62% in 2003. Many households, however, do not have multiple computers dedicated to each member living in the house. As such, sharing of computers amongst household members can be a prevalent phenomenon in home computer usage. Understanding this within-house computer sharing phenomenon and identifying the mostly likely person behind the computer screen can be of interest to market researchers and practitioners, particularly those interested in studying effective ways to target online ads based on users, online activities. For survey researchers who are attempting to recruit hard-to-reach individuals like teens and young adults, understanding of computer sharing could help establish contact at times when those individuals are more likely to be behind the computer. Despite its prevalence, within-house computer sharing has barely received any research attention. This study hopes to break through the barriers preventing the light of scientific inquiry into this phenomenon.

In this study, using survey data from a probability based online panel, we examine patterns of computer sharing among household members and their relationships with various online activities such as website visitations and social media access. We found that the depth of computer sharing, as measured by number of shared users in a home, manifests in different online behaviors at person and household levels. Further, as expected, depth of computer sharing was associated with higher concern for privacy in social media access. The study findings will likely have important implications for targeted online surveys, market research studies, and multi-level recruitment (household and then member) in online surveys. We discuss the findings and conclude with recommendations for future research.

## **Do Sequence and Mode of Contact Impact Response Rates for Web Only Surveys?**

Michael Stern, *NORC at the University of Chicago*

Ipek Bilgen, *NORC at the University of Chicago*

Kirk Wolter, *NORC at the University of Chicago*

The goal of many web based survey experiments is to drive households to the less-expensive web mode. However, many of the techniques for driving members of the general population using probability-based methods have not been as successful as possible due, in part, to the lack of viable, non-burdensome modes of contact. The significant issue that has limited research on probability-based web-only survey use in the general population is that email addresses are commonly unknown at the individual-level and, when they are available, they are mostly unmatched to other modes of contact such as physical addresses and telephone numbers. Due to the lack of readily accessible email addresses at the individual-level, surveyors are often forced to send postal mail recruitment letters, which provide a URL that the respondent must type into their browser. This extra burden results in lower response rates especially among those with low internet proficiency or those using a device other than a computer. Given this point, we need to consider ways of matching emails to physical addresses for the general population and testing multiple modes of contact. Using a general population

sample with matched email and physical addresses, we effectively test the mode of contact (email or URL embedded in an advanced letter), the sequence of contact (email or URL embedded in a letter first), and the efficacy of email versus mail follow-ups. Additionally, we evaluate the use of propensity weighting in internet samples. To date, we know of no research that has effectively achieved these goals among the general population. The results therefore help to shed light on the effects of mode and sequence of contact, incentive delivery, and post collection adjustments in web surveys.

## **Characteristics and Behaviors of Professional Respondents on Online Opt-In Panels**

Chan Zhang, *University of Michigan*

Christopher Antoun, *University of Michigan*

Huiying Yan, *University of Michigan*

Frederick G. Conrad, *University of Michigan*

Roger Tourangeau, *University of Michigan*

Mick P. Couper, *University of Michigan*

Professional respondents, those who complete large numbers of surveys, are one source of concern about data quality in online opt-in panels. Despite various claims about professional respondents, there is very limited published evidence regarding who the professionals are and whether they are indeed worse than other respondents. To investigate this, we analyze data from four independent surveys of opt-in panelists from 2008 to 2012, with about 2500 respondents in each. We classify as “professionals” those who belonged to 7 or more panels and had completed more than 30 Web surveys, “novices” as those who belonged to only one panel and had completed no more than 5 Web surveys and the rest as “average panelists”. Overall, a considerable proportion of panelists belonged to multiple panels, but many more than we expected were middle age or even older than 65. The likelihood of being professionals tends to be higher among older adults and lower among those with full-time jobs and higher income. Surprisingly, it was the novices not the professionals who exhibited more evidence of low quality data: novices were more likely to speed and straightline on grid questions than were professionals or average respondents. Contrary to the conventional wisdom, this finding suggests that, at least among online panelists, professionals are not necessarily bad respondents. The results also raise questions about what it means to be a professional respondent. It is possible that if professional panelists are really “professional” they may be motivated to do a reasonably good job to secure their jobs (i.e., not being kicked out of the panel). By contrast, novice panelists may be in the panel for a variety reasons and their behaviors can be more unpredictable.

## **Using Longitudinal Paradata to Explain Item Level Response Times in Online Panel Surveys**

Michael Bosnjak, *GESIS Leibniz Institute for the Social Sciences*

Mario Callegaro, *Google UK*

Kai Weyandt, *GESIS Leibniz Institute for the Social Sciences*

Survey item response times are widely been used to explore measurement quality and nonresponse in cross-sectional online surveys. We are however not aware of any attempt to connect such paradata across online panel waves in a longitudinal fashion. In the present research, we fill this gap by using survey response time paradata across the first eight survey waves of a probability-based online panel in Germany.

Our analyses focus on the time taken to answer a question as a function of item level characteristics (e.g. question type, content), respondent characteristics, survey topic, and the serial position of a survey wave. Using multilevel modelling approaches, we explore how these factors influence response times.

Preliminary analyses indicate that respondent characteristics seem to contribute most in explaining completion times. In contrast, the serial position of a survey wave does not serve as a significant determinant. We will derive actionable recommendations for online panel data quality management.

**Saturday, May 17**  
**10:00 a.m. – 11:30 p.m.**  
**Concurrent Session H**

## **Questionnaire Design: Topical Issues and Practical Applications**

### **The Effects of Question Order and Response Option Order on Self-Rated Health**

Dana Garbarski, *University of Wisconsin-Madison*  
Nora Cate Schaeffer, *University of Wisconsin-Madison*  
Jennifer Dykema, *University of Wisconsin-Madison*

The self-rated health question (“would you say your health in general is excellent, very good, good, fair, or poor?”) is widely used to study health because of its ability to predict morbidity and mortality. However, it is unclear from prior research how the distribution of answers to self-rated health (SRH) and their association with other health measures depend on (1) whether SRH is administered before or after more specific health-related items or (2) the order in which the response options are presented.

Keller and Ware (1996) recommend asking SRH before other questions about health and disease in order to remove the effect of context on respondents’ answers. However, preceding SRH with health-related questions may define and stabilize the meaning of “health” as a summarizing question. Results are mixed for the few studies that examine whether SRH’s distribution is affected by being placed before or after other health items. Further, no study examines whether the correlation between SRH and other health items changes depending on SRH’s placement.

Because research indicates respondents often select the first option they find acceptable, question writers often suggest beginning with the least desirable response option. However, SRH is overwhelmingly administered beginning with “excellent,” the most positively valenced option, and little research examines whether this ordering affects responses.

We analyze results from a 2-by-2 factorial experiment embedded in a web panel originally based on a national sample that varied the placement of SRH and the ordering of response options. We examine several outcomes including effects on response distributions and correlations with other health measures.

### **Making the Voice Heard: Added Value through the Use of Open-Ended Questions in the Context of Policy-Making**

Juliane Achatz, *Institute for Employment Research*  
Inna Becher, *Institute for Employment Research*

Social surveys often gain information which is used by policy-makers. Economic deprivation among children and adolescents is one such highly relevant issue in Germany. As children face an increasing risk of living in relative poverty, Germany increased its political efforts to prevent children from detrimental effects of growing up in impoverished conditions. In 2011 the “Education and Inclusion Subsidies Program” for children in low-income families was introduced.

In our paper we address the question, whether open-ended questions provide additional information which can be used successfully by policy-makers. More precisely, we investigate the response patterns of the respondents regarding their proposals for improvements of the “Education and Inclusion Subsidies Program”. We are particularly interested in whether the most disadvantaged social groups use open-ended questions and, thus, make their voice heard, or whether open-ended questions are mostly used by those social groups whose interests are already represented by the policy-makers.

Our analysis is based on the German Panel Study “Labor Market and Social Security” (PASS). The questionnaire includes several items concerning knowledge and usage of the “Education and Inclusion Subsidies Program” as well as open-ended questions regarding proposals for improvements of the program. First results indicate an unequal use of the open-ended questions. As such, persons in households with pre-school children use them significantly less. On the other side, the most disadvantaged households—those who receive basic income support—show significantly higher use of the open-ended questions and make various suggestions to improve the current subsidies’ program. Thus, our study makes a contribution, demonstrating the usefulness of open-ended questions in the framework of evaluating social policy measures.

### **Testing the Principles of Optimal Questionnaire Design: Does a Questionnaire Supposedly Designed Better Actually Work Better?**

Jon A. Krosnick, *Stanford University*

Soohee Kim, *Stanford University*

Russell Berman, *Stanford University*

The literature accumulated over the last 100 years points to a series of principles for designing questions optimally. If these principles are put into practice, do they actually improve the performance of the questionnaire? This study tested that notion beginning with the questionnaire used by a major American university to evaluate the quality of its teaching. The existing questionnaire violates many principles of optimal design, so two new questionnaires were designed according to best practices. Focus groups were conducted first to identify constructs to measure. Then questionnaires were drafted and cognitively pretested and revised. Finally, a field experiment was conducted where 900 students were randomly assigned to complete either the old questionnaire or one of the new questionnaires. The performance of the questionnaires were compared in terms of response rates, missing data rates, breakoff rates, completion time, and satisfaction ratings. The results not only identify a questionnaire that might be useful at a wide range of universities across the world but also reinforce confidence in the principles of optional questionnaire design.

### **Are Self-Assessments of Health Status Reliable Measure? An Examination of Potential Biases**

Mark Andrews, *Ipsos Health Policy Institute*

Self-assessment of health status is a standard question in health care research used in hundreds of thousands of interviews every year in quality assessments surveys and in countless other research instruments related to healthcare and a wide array of other subject matters. These measure are widely used and reported because they are easy to communicate and are have readily available benchmarks.



Those who use these measures, however, often do not question their validity. These questions are inherently subjective and are influenced by the relative expectations of the respondent and their disposition towards optimism.

This paper will examine this potential bias to determine how different groups respond to a standard self-assessment question. An experimental questionnaire will be designed and administered over the internet that will include this standard question and a number of more objective measures of health status. In addition, the respondent will be asked a series of questions assessing the respondent's disposition towards optimism. Analysis will examine if this optimism scale and/or a number of standard demographics explains divergence between the subjective and objective measures of health status. This paper hopes to provide insight into the magnitude of any measured bias, the characteristics of the bias, and how this bias may affect a commonly pointed to relationship between health status and overall satisfaction with one's health care.

### **Increasing Cross-National Comparability of Disability Measures with Anchoring Vignettes**

Mingnan Liu, *University of Michigan*

Sunghee Lee, *University of Michigan*

Mengyao Hu, *University of Michigan*

Colleen McClain, *University of Michigan*

While assessing well-being is an important topic for cross-national studies, uncertain comparability of its measurements is an impediment. One source of measurement incomparability is the differential response scale usage, which is due to either the cultural differences or the questionnaire translation. This error source produces less-than-perfect equivalence in a multilingual research setting, and renders estimated differences uninterruptable.

This study attempts to understand well-being across 11 countries through disability measured in five domains, namely pain, mobility, cognition, breathing, and affect, utilizing data collected in multiple surveys: the Health Retirement Survey, the Survey of Health, Ageing and Retirement in Europe, the English Longitudinal Study of Ageing, and the Chinese Health and Retirement Longitudinal Study. Respondents self-rated their disability on the scale of “none,” “mild,” “moderate,” “severe,” and “extreme.” This scale is comprised of vague quantifiers, and its usage is likely to be subject to cultural influences. These data sources also include a set of virtually identical anchoring vignette items, a popular tool developed to alleviate the influence of response scale usage differences in cross-national comparisons. For each disability domain, three vignette questions were asked in all these data sets.

This study will first compare disability in these countries solely based on the self-assessments. Then, we will also examine the assumptions of analyzing anchoring vignette data in these data sets. Third, we will use anchoring vignette data to correct for scale usage differences by comparing the differential item functioning-corrected self-assessments across countries. This is a nonparametric method. Last, we will use chopit model, a parametric method to correct for the cross-country differences of scale usage.

## Health Survey Research Methods I

### **Medicaid Reporting in the American Community Survey: Findings from Linked Administrative and Survey Data**

Kathleen T. Call, *University of Minnesota, SHADAC*  
Michel Boudreaux, *University of Minnesota, SHADAC*  
Joanna Turner, *University of Minnesota, SHADAC*  
Brett Fried, *University of Minnesota, SHADAC*  
Brett O'Hara, *U.S. Census Bureau*

Surveys provide critical data for evaluating the Patient Protection and Affordable Care Act (ACA). However, measuring health insurance coverage is challenging and virtually every survey is said to undercount Medicaid enrollment (i.e., survey counts of Medicaid participation are lower than is reflected enrollment data). We examine the extent to which Medicaid participation is misreported in the American Community Survey (ACS) among people known to be enrolled in Medicaid based on administrative records. We use a unique version of the 2008 ACS that has been linked to the 2008 monthly Medicaid Statistical Information System (MSIS) available in the Census Bureau's research data center located. We find that 77% of ACS respondents known to have Medicaid accurately report that they have Medicaid coverage. Of the 23% who misclassify their coverage, about 13% indicate another type of insurance and 10% report that they are uninsured. Consistent with research on other surveys, reporting of coverage is more accurate for children and individuals with lower incomes. The undercount varies across states from a high of 33.7% in Nevada to a low of 9.5% in Vermont. We examine the extent of the undercount across of variety of other characteristics potentially related to accurate reporting including household composition, native language, and mode of interview. We also discuss the role of the Medicaid undercount in the level of bias in aggregate estimates of coverage versus no coverage. We conclude that Medicaid reporting in the ACS is consistent with other federal surveys such as the National Health Interview Survey despite large differences in survey design. The results should provide policymakers confidence in using the ACS data to evaluate the impact of the ACA at the national and state level.

### **The Impact of Question Order on Familiarity with the Affordable Care Act (ACA): Part-Whole Order Effects and Health Policy Knowledge in Minnesota**

Alisha B. Simon, *Minnesota Department of Health, Health Economics Program*  
Kathleen T. Call, *University of Minnesota, SHADAC*  
Stefan Gildemeister, *Minnesota Department of Health, Health Economics Program*  
Susan Sherr, *SSRS*  
Giovann Alarcon, *University of Minnesota, SHADAC*  
Jessie K. Pintor, *University of Minnesota, SHADAC*

With major provisions of the Affordable Care Act (ACA) going into effect in January 2014, an assessment of public understanding of the law is essential to inform its ongoing implementation. To better understand familiarity with the ACA, two questions were added to the biennial Minnesota Health Access Survey (MNHA) in 2013. (The MNHA is a dual-frame RDD/cellular telephone survey, which provides estimates of health insurance coverage and access in Minnesota.) One question asked if the respondent felt that they knew enough about the health reform law to understand how it would affect them ("general impact question" adapted from the Kaiser Family Foundation). A second question set asked about familiarity with 5 specific provisions or "parts" of the law: Medicaid expansion, insurance exchanges, premium subsidies, no pre-existing conditions (guaranteed issue) and the individual mandate. After randomly

assigning the order of these two questions to all survey respondents, we test whether these questions are sensitive to part-whole order effects. We hypothesize that leading with the specific information provided by asking about certain provision of the law, some of which would apply to respondents, allows them to incorporate this information into their response to the general impact question. Preliminary results indicate that those who were asked the general impact question before hearing the questions about the specific provisions of the law were less likely to indicate they had enough information to understand how the law would impact them. Early results also find familiarity with different parts of the law was not appreciably different between the two conditions for all but the individual mandate, indicating that response of “yes” to the general impact question did not necessarily indicate increased familiarity with the law. Our preliminary findings indicate that part-whole order effects, whereby the parts inform the whole question, are present for knowledge questions.

### **Developing a Measure of Health Insurance Literacy**

Beth Forsyth, *American Institutes for Research*

Kathryn Paez, *American Institutes for Research*

HarmonieJoie Noel, *American Institutes for Research*

The Patient Protection and Affordable Care Act (ACA) has the potential to transform the health care access of more than 34 million previously uninsured, largely low- to moderate-income Americans. Fulfilling the potential of this significant legislation hinges upon consumers’ understanding of health insurance and their options for private health care coverage under the ACA. Little empirical information is available about what consumers in the private marketplace understand (Consumers Union et al, 2011; Loewenstein, Friedman, McGill et al, 2013). Data about what consumers know is critical to understanding how to communicate insurance concepts and benefit information effectively. To fill this void, the American Institutes for Research (AIR) developed a measure of health insurance literacy (HIL), a measure objectively assessing what consumers do and don’t understand as they select private health insurance and navigate the benefit structure of their plans.

We developed the draft HIL based on results from three research steps. Results will identify the different types of information available from each step. In step 1 key informant interviews detected two general sets of challenges related to (1) choosing and (2) using insurance. In step 2, consumer focus groups identified major knowledge, skills and behaviors required to select and use health insurance and reviewed draft items. Focus group analyses indicated four domains to measure: insurance knowledge, information seeking skills, general document literacy and cognitive skills. Step 3 tested revised items in two rounds of cognitive interviews. Revisions involved clarifying and simplifying question wordings and dropping questions that did not perform well. Research reported elsewhere conducted additional research assessing the revised the measure’s reliability and validity.

### **Trends in Attitudes Toward Health Insurance and Health Care and Their Influence on Health Care Behaviors**

Steven B. Cohen, *Agency for Healthcare Research and Quality*

Rising health care prices, the affordability of health care services and the ability to pay medical bills continue to stand out as major concerns when assessing public views on issues of health care costs and access. There is a growing body of evidence that demonstrates the influence of health care preferences on behaviors with respect to health insurance coverage decisions and health care expenditure behavior. Health insurance helps individuals receive timely access to

medical care and protects them against the risk of expensive and unanticipated medical events. In addition to the socio-economic profiles that distinguish individuals with coverage from those who are uninsured, attitudes regarding the need for and value of health insurance coverage may also have an impact on coverage decisions. Given the potential for these health care preferences to influence health behaviors, it is important to measure the population's attitudes towards health insurance coverage and examine the persistence of these attitudes over time. To support behavioral studies of health care coverage and expenditures over time, the Medical Expenditure Panel Survey also contains a series of self administered questions which discern individual attitudes regarding the need for health insurance coverage and its cost. In this study, we examine the attitudes adults have regarding the need for health insurance coverage and its cost. The study provides a detailed analysis of the level of concordance over time in health care opinions which assess the value of health insurance coverage based on self-reported data on these dimensions of individual health care preferences. The Affordable Care Act was enacted with major provisions to expand health insurance coverage and improve the health care delivery system. The analyses conducted in this study provide important insights regarding the significant influences exerted by health insurance preferences upon forthcoming coverage enrollment decisions and related interactions with the healthcare system.

### **Understanding Attitudes Towards “Obamacare”: Question Wording and Terminology Usage in Public Opinion Surveys on the Affordable Care Act**

Megan Moldenhauer, *Ipsos*

Chris Jackson, *Ipsos*

Julia Clark, *Ipsos*

John P. Vidmar, *Ipsos*

As arguably the most notable piece of legislation passed by the current Administration, attitudes towards the Affordable Care Act (ACA) have been much-researched by pollsters, academics, and politicians alike. Our research question is to determine to what extent political values that have come to be identified with the ACA have overridden opinions of the law itself? The term “Obamacare” was originally utilized as a pejorative to attack the ACA, but is now regularly used, and globally recognized, by both supporters and detractors. In the process, the term has become a lightning rod, encompassing broad opinions, political influences, and policy attitudes, rather than an identification of policy specifics. Increasingly, we are seeing the possibility of a divergence between opinions of “Obamacare” (interpreted as a concept rather than a policy) and opinions of the ACA.

Through the use of original ongoing research utilizing the Reuters/Ipsos poll, this paper seeks to understand how the deployment of different terminologies language may impact measurements of public attitudes towards the ACA. We do so through the use of a split ballot experimental design with a representative survey sample of over 2,500 Americans. We evaluate the term “Obamacare” along with additional terms relating to the Exchanges and access points, and policy-based language. Where possible, we seek to understand how the substitution of language used commonly by the media (as compared to the legislative language) might sway the public in terms of both familiarity with and favorability towards various aspects of the Affordable Care Act.

Understanding the impact of different question iterations is critical to assessing where true public opinion on this issue lies. Truly accurate measurement of public opinion plays a critical role in a democratic nation, and this piece of legislation is likely to be the one that defines President Obama's legacy.

## **Election Polling Methods**

### **States as a Natural Experiment: Survey Mode Biases in the 2008 Exit Poll**

Ozan Kuru, *University of Michigan-Ann Arbor*

Josh Pasek, *University of Michigan-Ann Arbor*

Exit polls provide an important window into who votes in American presidential elections. But effective exit polls can no longer rely entirely on exit polling. With the emergence of early voting and vote-by-mail precincts, telephone interviews have accompanied election precinct surveys to capture voter demographics and preferences. This structural change has introduced the potential that our understanding of the dynamics of elections may vary depending on which survey mode was used (in person or telephone). Because the methods used for the exit poll depend on the prevalence of early and absentee voting in a state, these dynamics may undermine attempts to compare the electorate across states. To explore the role of survey mode in altering our understanding of the electorate, this study treats variations in state election laws and the changing design of the exit polls as a natural quasi-experiment. Results from the exit polls in each state were compared with data from the Current Population Survey's voting supplement, which used a consistent mode for all 50 states. Differences emerged between states that used each of the two modes; notably, young voters were represented at much higher rates in states without early and mail-in voting. Disentangling the phone and in-person portions of the exit poll indicated that this was the product of a survey mode effect. Unfortunately, freely available exit poll data does not provide sufficient information to re-weight phone and in-person portions of the survey to account for this bias.

### **Exit Polling vs. Telephone Survey Research: A Methodological Test**

Fred Solop, *Northern Arizona University*

Emily Schnurr, *Northern Arizona University*

Exit Polling is widely considered to be a gold standard for understanding voter preferences. But, in today's world of early voting, permanent early voting, and no excuse voting, exit polling methodology needs to be supplemented with telephone survey research to account for voters participating in an election before election day. While this is becoming regular practice within the polling community, the nuances of this mixed methods approach has yet to be discussed in the literature.

Flagstaff, a mountain community in northern Arizona, asked voters in November 2012 if they support or oppose bonding to clean the local watershed via a partnership agreement with the U.S. Forest Service. Voters overwhelmingly approved the measure. Anticipating the historical significance of an environmental measure of this type, researchers at Northern Arizona University sought to understand voter motivations. We conducted both an Election Day exit survey and a post-election telephone survey of early voters in pursuit of this question.

This paper begins with the data and compares results achieved from the exit poll and the phone survey. The central question is whether the results internally valid, or whether data findings are actually influenced by the methodology of data collection? The paper goes on to discuss the learning curve researchers experienced in blending the two methodologies together.

## **Treatment of Refusals in Exit Polls**

Vadim Volos, *GfK*

Respondent refusals in exit polls present a serious challenge to pollsters, impacting both estimation of total error and final projection of election results. In exit polls conducted by various research organizations across the world in the past years, refusal rates stay consistently high and it seems no silver bullet has been found yet to treat them extremely effectively. Pollsters should continue to pursue better understanding of where refusals are coming from and how to best minimize them, as well as how to treat them in post-survey estimation and analysis.

In this paper, I draw on literature and experience of exit polling by GfK and other research firms to explore the following:

- Sources and root causes of refusals in exit polls
- Different types of refusals
- Control of refusals by research organization and design
- Statistical treatment of refusals

How this all works in a specific exit poll? I will talk about our battle experience with a recent exit poll on the Presidential elections in the former Soviet republic of Georgia, held on October 27th, 2013 and present our team's approach to refusals in this exit poll and end results.

## **Constructing a Likely Voter Model with Weak Priors: Evidence from 3 Massachusetts Elections**

Joshua Dyck, *University of Massachusetts Lowell*

Francis Talty, *University of Massachusetts Lowell*

David Barney, *University of Massachusetts Lowell*

Getting an accurate likely voter sample when using RDD can be challenging due in large part to a strong social desirability bias in over-reporting the propensity to vote among respondents. This is complicated by the fact that political campaigns may directly target and mobilize voters with a lower probability of voting. The approach to these situations relies, in part, on answers to a series of questions meant to reveal a respondent's intention to vote, interest in politics, and knowledge about the upcoming election. However, getting to an accurate LV sample also relies on strong priors of turnout from previous elections. Occasionally, the political arena presents researchers with atypical elections for which we lack information that allows us to reliably predict turnout with much precision. While such elections tend to crop up sporadically, the state of Massachusetts has recently gone through a series of irregular, off-cycle elections –special elections for both US Senate seats in a period of three years, followed by the first open-seat-competitive Boston Mayoral election in two decades. Analysis of registered voter samples from Massachusetts and Boston over a series of elections, collected using a dual-frame RDD approach out of which likely voters were sampled, provides us with leverage in attempting to unpack the precise estimate of likely voter populations in special elections. We compare the effectiveness of different likely voter models and conclude with recommendations for future special election survey research.

## **The Booker Effect: The Role of Interviewer Race and Ethnicity in Polling the 2013 New Jersey Special U.S. Senate Election**

David Redlawsk, *Rutgers University*

Ashley Konig, *Rutgers University*

Cory Booker (D), the African-American Newark, NJ Mayor known for his tweeting and headline-making, was expected to beat Steve Lonegan (R) in a special U.S. Senate election in October. As the campaign progressed, multiple polls showed Lonegan closing in on Booker – most likely due to Booker’s lackluster campaign – but the results of each poll varied significantly due to the difficulty predicting likely voters in such circumstances. But at the end, most polls converged on similar estimates – except for the Rutgers-Eagleton Poll, which showed a 22-point margin for Booker, well above other polls, and which proved to be more than twice Booker’s actual margin.

The root of this discrepancy may stem from a Rutgers-Eagleton Poll “house effect” due to significant demographic diversity among its interviewers, drawn from the highly diverse Rutgers University student population. Recent Rutgers entering classes have been majority-minority, reflected in the fact that call center staff is only about 25% white.

Analysis suggests Rutgers-Eagleton suffered from a “race of interviewer” effect, where the diversity among the poll’s callers prompted a greater propensity for respondents to report a Booker vote. Specifically, analyses find a significant main effect for interviewer race, where non-white interviewers were more likely to record Booker votes than whites, even after controlling for classic vote predictors, respondent demographics, and interaction between caller and respondent race. This effect was absent in data collected at the same time on the NJ gubernatorial race, which featured two white candidates, and where Rutgers-Eagleton reported results similar to other polls. The greater accuracy of other polls on Booker may be due to less diverse call centers or the use of IVR data collection.

Results suggest that race of interviewer effects are alive and well, having implications for African-American candidates, live interviewing, social desirability bias, and telephone survey interviewer populations.

## **Big Data in Public Opinion and Survey Research**

### **How Can “Big Data” be the “Data” for Survey and Public Opinion Researchers?**

Kumar Rao, *The Nielsen Company*

The term Big Data (BD) is the “buzzword du jour.” It is often invoked to describe the overwhelming volume of information produced by and about human activity, made possible by the growing ubiquity of mobile devices, tracking tools, always-on sensors, and cheap computing storage (Lewis, Zamith, & Hermida, 2013). How is BD different from any other data? As Sathi (2013) notes, for data to be referred to as BD it is characterized by four elements: volume, variety (support for different data types), velocity (speed of data processing and computation), and veracity (credibility and suitability of the data source).

The question that confronts survey and public opinion researchers is does BD have any relevance and significance in their line of work? Is it possible to best leverage the traditional capabilities and principles of data collection while taking advantage of the efficiencies of computational methods and technologies offered by the BD phenomenon? This has been a

subject of discussions and debate, with proponents and opponents pointing toward differing implications. Some researchers have argued that social science, which traditionally has relied on relatively small samples using tools such as surveys, can make use of large quantities of real-time transactional data on the underlying population of interest. This, they argue, would enable researchers to become much more powerful in answering certain types of questions. On the other side, researchers argue small-scale analyses can provide equally meaningful inferences when using proper sampling, measurement, and analytical procedures. They argue whether analysis of huge quantities of data is theoretically justified given that it may be limited in validity and scope. AAPOR, as an independent, non-agenda driven, community-centric organization, finds itself in the center of this debate. This presentation will introduce the BD phenomenon and the considerations for its use in public opinion research.

### **Towards a Total Error Framework for Big Data**

Paul Biemer, *RTI International*

A framework for describing the total error of a data set or estimator is well-established in the survey literature. This framework or paradigm underpins best practices regarding survey design, implementation, evaluation, and analysis for ensuring maximum accuracy in the survey results under fixed budget constraints. There is tremendous interest in the survey community to incorporate big data – i.e., social media, transaction, sensor, administrative data, etc. – into survey estimation processes, so a total error framework that mimics the total survey error paradigm is needed. This is particularly important when big data are used to generate population estimates that either augment or supplant survey estimates. This paper considers the potential error sources and risks from using big data for these purposes and presents an error framework applicable to big data in the production of population statistics.

### **Big Data in the Census Center for Administrative Records Research and Applications**

Amy O'Hara, *U.S. Census Bureau*

The Center for Administrative Records Research and Applications (CARRA) processes and analyzes federal, state, and commercial data to improve operations and data products. Many of our files are large, but they are not necessarily "big data." We can improve our methods on these files using Big Data hardware, software, and methods. Other files really are Big Data, and we need to determine the quality and utility of these data for various purposes. For example, the data might be useful when developing a frame or modeling, but lack precision for allocation/item replacement purposes.

We are exploring the benefits of web scraping, improved parallel processing, and modeling. We use data that we have collected, often integrated with administrative records and third party data. We are identifying and addressing the risks of these approaches with regard to privacy and perception.

In this presentation, I will discuss the current and future work concerning the agency's approach/metrics around Big Data. I will also touch on the challenging statistical and consent issues.



## **Experimental Research, Machine Learning and the Next Generation of Microtargeting Models**

Andrew Therriault, *Greenberg Quinlan Rosner Research*

The past decade of political campaigning has been shaped by the rise of microtargeting, which combines surveys with voter and consumer data to create individualized "scores" for each voter. These scores have allowed campaigns to increase their effectiveness by targeting their outreach based on estimates of what voters think about issues, how they are likely to vote, and whether they will actually show up on election day. But for a data-minded campaign, knowing where voters stand now is only part of a bigger picture. The real prize? Knowing which voters can be moved, and how. During the 2012 elections, a handful of campaigns began to apply sophisticated predictive modeling techniques to data from experimental tests, using the results to successfully predict voters' responses to campaign appeals. As these tools are expanded and improved over the coming years, campaigns will be able to target individuals based on their receptiveness to persuasion and mobilization---and moreover, they will do so with messages tailored to the individuals' own characteristics and delivered through the most effective medium for each target. These advances will reshape how we experience and interact with political campaigns, and as a practitioner deeply involved in the creation of these tools, I will illustrate how they are developed, what they can tell us, and what they can offer to the field of opinion research more broadly.

## **Wearable Technology and the "Internet of Things": How These Big Data Will Change Survey Research**

Brian Head, *RTI International*

Over the past few years, survey researchers have become interested in how they can use big data to improve or supplement traditional survey methodologies. The focus on big data has primarily been reactive, focusing on devices and social media after they gain popularity. However, emerging technologies will produce exponentially more big data than current sources that have already achieved popularity. These technologies will provide researchers even more options to supplement surveys and help answer important social questions.

Two such technologies are wearable technologies and the internet of things (IoT). Wearable technologies are devices or sensors that have some utility. For example, the Nike+ Fuelband and FitBit are devices users wear to record biofeedback data. The IoT refers to an internet-like structure in which machines---physical and virtual---are connected and share data with each other. A close relationship between these two technologies already exists and will continue to grow in the near future. For example, wearable technologies like FitBit use the IoT to communicate with mobile applications, such as MyFitnessPal, to allow users to record caloric intake and output to guide weight loss or management. Data for such applications are often stored in the cloud---a database located on a server accessed over the internet. Such passively collected data are potentially richer and more accurate than retrospective self-report data because the response process, during which error can enter in multiple places, is removed from data collection. This presentation will examine the current capabilities of these emerging technologies, where they might be heading, and the potential for researchers to incorporate them into behavioral and social research.

## Social Media Data and Analysis

### **Best Practices for Using Phone, Email and Text Messaging to Increase Panel Participation**

Kelly Dixon, *The Nielsen Company*  
Eku Kendall, *The Nielsen Company*

Researchers today have a wide array of new and traditional communication tools to encourage daily participation in study panels. The question is: what is the right mix? This research demonstrates how a mix of communications modes including text messages, emails, pre-recorded phone messages and live web chat -- as well as continued use of mail and phone contacts -- can be used to effectively encourage panelists to comply with some fairly complex data collection tasks in a long-term panel. This presentation will describe and quantify the types of messages we utilize during the panelist lifecycle and discuss best practices for their use in maintaining panel participation and supporting performance improvement. Using data from an on-going panel measuring media consumption, we show that a mix of communications makes it more likely that we reach panelist with diverse preferences and allows us to deliver multiple reminder messages without it seeming overly repetitious to the participants. Text messages are particular useful at reaching key demographic sub-groups that have mobile devices and are heavy users of this type of communication. The right mix also increases cost-effectiveness by targeting low-cost communications (such as texts, emails and automated phone messages) where they are effective and reserving higher cost modes (live telephone or in-person coaching) where they are most needed.

### **Surveys of Facebook Users vs. Integrating Facebook in a Population Based Survey: Enhancing Survey Research on Social Networks**

Stuart Michaels, *NORC at the University of Chicago*  
Patrick van Kessel, *NORC at the University of Chicago*  
Michael Jugovich, *NORC at the University of Chicago*  
John A. Schneider, *University of Chicago*

Social media platforms are increasingly being used in social science survey research, with particular interest in Facebook due to its popularity and large active online community. Recent studies have used custom Facebook applications to conduct surveys of users of the social media website, combining their responses with data about their Facebook activity. However, while these studies have yielded intriguing results that correlate online behaviors such as Facebook “Likes” with survey measures of demographic characteristics, personality traits, and intelligence, their findings apply exclusively to self-selected active Facebook users.

This presentation will describe an alternative approach taken in an ongoing Respondent Driven Sample survey of a classic hard-to-reach population, young Black men who have sex with men. Facebook data are integrated directly into the survey to augment and enhance data collection efforts without influencing sample design. In contrast to surveys deployed directly on Facebook, a custom Facebook application is embedded into an independently administered CAPI survey, allowing researchers to augment data collected in the survey by dynamically modifying the survey itself in real-time through the use of an algorithm that ranks and selects a set of Facebook friends from each respondent and generates subsequent survey questions about the resulting social network roster.

After describing the innovative aspects of this methodology, we will explore the unique analytic utility and limitations of the resulting data through a comparison of Facebook and non-Facebook respondents and an analysis of survey responses in relation to the social network structure derived organically from Facebook. We will then conclude with remarks about the broader scientific compatibility of survey and social media research methods, and consider the implications of the described method and other potential applications for the future.

## **Can Social Media Data Predict Survey Data? A Meta-Analytic Review of the Literature**

Lauren Guggenheim, *University of Michigan*

Josh Pasek, *University of Michigan*

Cliff Lampe, *University of Michigan*

Michael F. Schober, *New School for Social Research*

Frederick G. Conrad, *University of Michigan*

Ellen Wagner, *University of Michigan*

Lindsay K. Brown, *University of Michigan*

Increasingly, researchers are exploring when big data analytics using social media data produce trends that track survey results. A number of recent studies have found that machine learning algorithms and sentiment analysis can predict changes in consumer attitudes, political preferences, and public opinion. Yet it remains unclear when analyses of big data will reproduce the results of opinion surveys and when the trends will diverge. To understand the parameters of current research, we conducted a comprehensive search of the relevant scientific literature and classified the studies. We identified substantive differences in the ability for social media data to match survey data depending on the topic (politics, consumer attitudes, etc.), the methods used (sentiment, mentions, etc.), and the presence of earlier studies in an area. These discrepancies imply that social media data may not be equally predictive across all arenas. Based on the literature reviewed, we propose when these data sources might be expected to diverge based on both the nature of the research questions and the community of users (survey respondents in different kinds of surveys vs. social media contributors with differing levels of activity and in social media with different properties).

## **When Are Big Data Methods Trustworthy for Social Measurement?**

Cliff Lampe, *University of Michigan*

Josh Pasek, *University of Michigan*

Lauren Guggenheim, *University of Michigan*

Frederick G. Conrad, *University of Michigan*

Michael F. Schober, *New School for Social Research*

A number of recent demonstrations in a variety of fields have shown that “Big Data” derived from social media sites may be able to measure phenomena traditionally derived from surveys (e.g. perceptions of well-being, political opinions, and consumer confidence) at much lower cost and without burdening members of the public. In this paper we articulate a set of distinctions for arguing on logical and theoretical grounds when big social data might augment or even replace survey measurement. These distinctions also lay the ground for empirical tests that will allow researchers to ascertain when alternative measurement methods can be trustworthy. As we see it, the major areas of difference between survey and Big Data from social media lie in (a) how participants understand the activity they are engaged in (responding to a survey or posting in social media), (b) the nature of the data to be analyzed, and (c) practical and ethical considerations surrounding use of the data. For (a), participants in social media have different

notions than survey respondents about why they are participating, what they have consented to, the extent to which they have previously thought about the topics under investigation, the representation of their identity, and the nature and size of their perceived audience. For (b), the data under analysis vary in how they represent the population, how they are structured, and in the very questions that they can answer; as we see it the central question is whether the sheer amount of data available from social media streams outweighs questions about representativeness of the users. For (c), there are substantial differences in costs to researchers, ownership of the data, potential stability of the data, and the depth of knowledge about appropriate methods in the relevant research communities.

## **Campaigning in the Digital City: Simultaneous Survey and Digital Analysis of NYC Mayoral Primary**

Micheline Blum, *Baruch College School of Public Affairs, CUNY*

Douglas Muzzio, *Baruch College, CUNY*

Joshua Hendler, *Hill & Knowlton Strategies*

Jacqueline Fortin, *Baruch College, CUNY*

Nicole Lee, *Baruch College, CUNY*

Amber Ott, *Hill & Knowlton Strategies*

Baruch College Survey Research and Hill+Knowlton Strategies collaborated on a public opinion poll of New York City designed to explore how New Yorkers use social media to engage with mayoral candidates and to learn about political news and simultaneously monitored social media and created a Digital Engagement Index ranking the NYC mayoral candidates' efforts on Facebook, Twitter, YouTube and through email.

Baruch conducted a telephone poll of a random sample of 1001 NYC adults in August 2013, including 691 landline and 310 cell phone interviews. Respondents were asked through which media they got their news, if they went online for any political news, which social media outlets they used for political information, if and through which media the various campaigns had reached out to them and the ways in which they were or were not engaged in the campaign.

The H+K digital analysis monitored the actual use of old and new media by the various campaigns in July and August 2013 and ranked their digital outreach.

The paper presents the findings of the two simultaneous studies and suggests ways of measuring and comparing campaigns through old and new media and methods.

## **Is There Still a Role for Credible, Non-Partisan POR in the Age of Sound Bites, Social Media and Politically-Driven Polling?**

### **Academic Perspective**

Robert Y. Shapiro, *Columbia University*

My discussion will address the questions posed for the panel. It will also emphasize how, in the same spirit, the role of public opinion research extends beyond this to tracking an increasingly important and visible aspect of our nation's political and social history and, further, to assessing the "democratic-ness" of American democracy. What matters here as well is the continued high quality of opinion polling and the access to the data provided by polling organizations and

survey sponsors--through their reporting of their research, making their methods transparent, and their willingness to archive data for further study and as part of the historical record.

### **Media Perspective**

Marjorie Connelly, *The New York Times*

More than ever, the public turns on the radio, television, or computer and tunes into an echo chamber reflecting back information filtered to their own way of thinking. Just as congressional districts have been gerrymandered into partisan strongholds, so have media and internet outlets. Increasingly, Americans read and listen to people with beliefs similar to their own, those who are aligned with their own ideology. Legitimate, scientific polls with carefully worded questions reach out and get at what Americans truly think about the important issues facing the country -- with no political bias. Polling results this year have showed that Americans held opinions about healthcare, gay marriage, immigration, and gun control laws that were not always in lock-step with their party identification or their political representatives. Without legitimate polling, the media would be forced to rely on "man in the street" interviews, suspect data from campaigns, and metrics from organized letter-writing operations and the social media equivalent.

The New York Times and other media organizations must continue to locate and utilize unbiased, representative and reliable public opinion research. Reporting survey research is no different from reporting any other information. When inferior research is used, not only have the viewers and readers been misled, but the credibility of the media has been hurt. As long as quality survey research exists, there will be a role for it in both traditional and new media outlets. Credible public opinion research provides a reality check for decision-makers and political leaders, who themselves are inundated by what's trending in the mainstream media, or on their Twitter feeds. Legitimate polls are able to speak "truth to power," providing a means of critical discourse within the echo chambers that threaten to drown out American voices.

### **Non-Profit Perspective**

Scott Keeter, *Pew Research Center*

V. O. Key, Jr. defined public opinion as "those opinions held by private persons which governments find it prudent to heed." One premise of this panel is that there is public opinion that governments do not always heed, or even aware of. It is the mission of public pollsters to gauge public opinion, primarily through survey research, and make it widely available so that policymakers and other influentials at least have the option to heed it. Two trends are worth noting for their potential impact on the ability of public pollsters to achieve this mission. One encompasses financial and other pressures that are making it more difficult for public polling organizations to function effectively. These include the ongoing retrenchment in traditional news organizations that sponsor polling, and similar challenges in academia that affect university survey centers. This retrenchment has the effect of reducing the volume of polling and the number of staff available to interpret public opinion, whether generated by the organization or by others.

The second trend is the growth of alternatives to polling as a source of data on public opinion. This trend may have both positive and negative effects on the ability of organizations to describe public opinion. As Robert Groves has written, the future of survey research may entail an integration of "designed data" -- as collected by the traditional sample survey -- and "organic data" -- from non-survey sources such as financial transactions or social media. However, the use of organic data to represent public opinion is still in its infancy and needs careful study and

evaluation before it can be trusted in the same way that random sample survey data is trusted. The field needs successful examples of how designed and organic data can be combined to yield greater insights.

### **Commercial Perspective**

Nancy Belden, *Belden Russonello Strategists*

Credible POR conducted by for-profit firms continues to be in demand by client organizations and elected official who want to influence public opinion and influence policy making. The clients also may utilize non-survey sources such as Twitter and big data aggregators, or poor survey sources such as IVR and Survey Monkey in the hands of amateurs. But because advocacy groups and politicians want to understand issues or campaigns fully and honestly so they can communicate well and win the day, there is a great deal of demand for sophisticated survey research that goes on behind the scenes and the public will never see. Good policy makers are using this research to understand voter and general public opinion and to learn how to try to get the public to understand why they advocate a policy -- and perhaps to move the public to support their policy initiatives. Also, surveys released by the public pollsters are important as a general guide and historical record. Some of that research is deep but much of it skims the surface, has its own bias, and can be misleading.

## **Testing New Methods for the 2020 Census**

### **Exploring New Methods for Census Data Collection: The 2013 Census Test**

Peter V. Miller, *U.S. Census Bureau*

A key goal of research in preparation for the 2020 Census is to identify ways to contain costs. Enumerating people who do not self-respond presents the main challenge to cost containment. In previous Censuses, non-responding households have been contacted in person. This effort has involved a very large and costly field operation. The 2013 Census Test is the first study to examine potentially more efficient methods to enumerate those who do not self-respond.

The test is an exploratory study of administrative records use, alternative contact strategies and mode alternatives. First, it seeks to examine how administrative records might be used to reduce the nonresponse follow-up (NRFU) workload, or to aid in prioritizing field effort. Second, it examines two ways to manage the interview workload – a “fixed” and an “adaptive” approach. Third, it examines two ways to execute telephone contacts – through a call center or through dispersed individual calling.

A sample of approximately 2000 addresses in Philadelphia whose occupants did not self-respond to the 2010 Census was selected. These addresses were randomly assigned to different experimental treatments that use administrative records in different ways to aid in enumeration, employ “fixed” or “adaptive” field procedures and make use of telephone contacts either through a centralized CATI operation or through dispersed individual enumerator calls. Approximately 20 current Census field representatives were employed for fieldwork. The interviewers were divided into two groups, one to work in the “adaptive” contact condition and one to work in the “fixed” condition. The study was fielded in November and December of 2013.

This paper discusses the cost and data quality outcomes of the different methods examined in the test and the implications for future 2020 Census methodological research.

## **Using Administrative Records in the 2013 Census Test**

Andrew Keller, *U.S. Census Bureau*

Scott M. Konicki, *U.S. Census Bureau*

Tom Mule, *U.S. Census Bureau*

This paper documents and analyzes the implementation of administrative records usage in the 2013 Census test. The 2013 Census Test was a field implementation of adaptive design strategies for Nonresponse Follow-up (NRFU) contact. The 2013 Test consisted of an initial postcard mailing and subsequent NRFU visits for selected units. This test used administrative records in two ways. The first was to identify housing units that could be removed from the Nonresponse Follow-up workload. These removed units were determined to be either occupied or vacant based on the administrative records available and undeliverable-as-addressed information from the US Postal Service following the postcard mailing. The second main area was to provide to the adaptive design staff information about the household composition of the unit based on administrative records. These indicators were made available to help determine the adaptive strategy implemented in the test.

In this paper, we review the administrative records files that were available and utilized. The paper details the process by which the available administrative records were determined to be a suitable for the housing unit. Since some panels conducted fieldwork instead of removing cases from the workload, we provide analysis of the administrative records information to field interview results where available.

## **Interviewer Implementation of Adaptive and Fixed Data Collection Approaches in the 2013 Census Test**

Gina K. Walejko, *U.S. Census Bureau*

A central element of the 2013 Census Test was to examine the performance of adaptive and fixed data collection procedures in nonresponse follow-up. In the adaptive condition, personal interview contacts were guided by a centralized system that utilized frame and contact history information to prioritize cases for interviewers on a daily basis. In the fixed condition, interviewers were provided with sample cases at the beginning of the field period and given a global instruction to complete interviews efficiently. Both groups were allowed a maximum of three in-person contact attempts per case. In the adaptive condition, however, the number of contacts permitted varied by whether the Census Bureau had administrative records that could enumerate the household. If records were available, only one contact was allowed. In the adaptive condition, in-person interviewing was preceded by telephone attempts through a centralized CATI facility. In the fixed treatment, field interviewers were instructed to individually make two telephone contact attempts before any in-person attempts. In both conditions, interviewers were instructed to attempt a proxy interview if the sample household did not yield an interview on the final contact attempt.

This paper examines how these complex adaptive and fixed data collection procedures – unprecedented in Census data collection -- were implemented in the field. We measure how well CAPI interviewers followed instructions such as contacting specific households on particular days in the adaptive treatment, making telephone calls before personal visits in the fixed condition, completing cases in one or three contact attempts in the adaptive condition, and performing proxy interviews if needed on the last contact attempt in both conditions. We also attempt to measure how faithfully interviewers provide paradata on contacts and costs.

## **An Examination of Telephone Interviewing in the 2013 Census Test**

Gianna S. Dusch, *U.S. Census Bureau*

Sandra Ehni, *U.S. Census Bureau*

Caleb M. Miller, *U.S. Census Bureau*

As part of the 2013 Census test, telephone interviewing, which has not been widely used for Census nonresponse follow-up, was employed in two different ways. Telephone numbers from administrative records were matched to sampled addresses. For cases with matching telephone numbers, we attempted contact with the sampled housing units via telephone before any in-person contacts. For one treatment group, cases containing telephone numbers were sent to a centralized computer assisted telephone interviewing (CATI) center prior to being assigned to personal visit interviewers. For the second treatment group, decentralized interviewers in the field were instructed to first attempt contact via telephone when telephone numbers were provided, before attempting a personal visit.

This paper examines the telephone component of the 2013 Census Test. We review the process for obtaining and matching telephone numbers to addresses, as well as how numbers were prioritized for inclusion in the 2013 Census Test. The final disposition of the case is used to assess the quality of the telephone number. Additionally, we evaluate the two approaches for telephone interviewing, through a centralized calling center or through calls by decentralized field interviewers. Metrics such as rate of completion by telephone and date of completion will be examined.

## **The Civil Rights Act of 1964 and Public Attitudes on Race, Ethnicity and Gender: Fifty years later, What has Changed? What Remains the Same?**

Panelists:

James S. Jackson, *University of Michigan, Institute for Social Research*

Taeku Lee, *University of California-Berkley*

Ange-Marie Hancock, *University of Southern California*

Mark H. Lopez, *Pew Research Center*

This year marks the 50th anniversary of the Civil Rights Act of 1964, the most significant piece of social legislation passed in the last half of the 20th Century. Speaking from their areas of expertise, this distinguished panel of researchers presents what we know about the trajectory of public attitudes on race, ethnicity, gender and class from the 1960s to the present.



**Saturday, May 17**  
**12:45 p.m. – 1:45 p.m.**  
**Poster Session 3**

**1. Exploring the Use of Three Unmoderated, Online Testing Services for Evaluating a Survey Brochure**

William Mockovak, *Bureau of Labor Statistics*

Brochures can play an important role in survey response, and their effectiveness has often been evaluated using expert reviews, focus groups, or respondent debriefings. Although helpful, these approaches are constrained by small sample sizes, limited sample representativeness, and operational challenges. The current study explored the use of three online services – SurveyMonkey, Amazon’s Mechanical Turk, and TryMyUI – for obtaining feedback about a 14-page brochure designed to encourage survey participation and to increase awareness of statistical products.

Four pages from the brochure were evaluated by audiences recruited through SurveyMonkey. Each page had an independent sample, with sample sizes ranging from 71 to 367. The only selection criterion for a sample was age, which was allowed to vary between 18 to 65, depending on the sample. For two of the brochure pages, samples (N=102 and 98) were also obtained through Amazon’s Mechanical Turk to see if comparable findings resulted. Since respondent motivation was a concern, the online questionnaire was kept short (7-9 pages, with a maximum of 13 questions), and a small experimental study was embedded to determine if use of a progress indicator had any impact on data quality.

Since neither SurveyMonkey nor Mechanical Turk was appropriate for obtaining reactions to the entire 14-page brochure, we also recruited evaluators from TryMyUI, an online testing service that is geared toward usability testing, and which provides videos lasting up to 20 minutes of participants completing assigned tasks. Using TryMyUI, 22 participants reviewed the entire brochure, with 14 also giving detailed feedback on a single page of interest.

The three online approaches were easy to use and produced general agreement on broad measures of the brochure’s effectiveness, but the Mechanical Turk sample produced significantly more positive responses. The use of progress indicators had no effects on item completion or data quality.

**2. Differences in Early and Late Respondents: Comparing RDD and ABS Telephone Designs**

Brian F. Head, *RTI International*

David Roe, *RTI International*

Bonnie Shook-Sa, *RTI International*

Barbara Bibb, *RTI International*

Doug Currivan, *RTI International*

Empirical evidence suggests that late survey respondents are often significantly different on both demographic and key substantive measures from early respondents. Substantial

resources are often devoted to follow-up contacts in an effort to minimize nonresponse error. To date, research on minimizing nonresponse error in telephone surveys has primarily focused on Random Digit Dial (RDD) designs in which contact is generally made through outbound calls. Because of the challenges associated with geographically targeting dual-frame RDD samples, address-based sampling (ABS) has emerged as an alternative design for studies seeking to make sub-state estimates.

Data from two rounds of the Aligning Forces for Quality (AF4Q) Consumer Survey (AF4Q 2.1 and AF4Q 2.2) are used to compare early and late respondents on demographic and health measures. The target population for the AF4Q study consists of chronically ill consumers of healthcare. AF4Q 2.1 was based on a dual-frame RDD design, while AF4Q 2.2 utilized an ABS design in which phone numbers were reverse matched to addresses for approximately half of sampled households (matched). Those for which no phone number could be matched (unmatched) were sent a letter briefly describing the study, what participation meant, and a household information form. ABS unmatched households were asked to either call in to complete the study or return the household information form to be contacted by the research team. We compare the characteristics of early and late respondents across the two designs as well as between matched and unmatched case statuses. Findings from this study will be useful to researchers making decisions about the length of data collection and the level of effort for telephone studies when utilizing dual-frame RDD or ABS designs.

### **3. Please Hang Up and Dial Again: Examining the Impact of Increased Call Attempts on Telephone Survey Outcomes**

Daniel Bausch, *APPRISE*

Kathi Barringer, *APPRISE*

A primary concern for survey research practitioners is minimizing nonresponse bias that results from nonrandom differences between survey participants and non-participants. Researchers attempt to minimize nonresponse bias by employing techniques to gain cooperation from hard-to-reach or uncooperative sampled cases, including repeated contact efforts, refusal conversion attempts, and incentives. Yet, while these efforts have been shown to increase response rates, the existing research literature has found varying effects on the impacts of each effort on nonresponse bias. In addition, these efforts increase the length of the survey field period and the cost of administration. One challenge for all practitioners is selecting a survey design that balances nonresponse bias and administration cost.

The focus of this analysis will be estimating the impact of increased telephone survey call attempts on survey outcomes using survey responses and call-record data from a telephone survey. Some recent research has accepted response rates as low as five percent, and it is important to document that expending extra effort to increase the response rate yields higher quality data. This analysis will use data from a random sample of 4,000 utility customers in Northeastern United States who were selected for a survey about compact fluorescent light bulb (CFL) awareness and purchase behavior. Sampled cases received up to 21 contacts, and the final response rate was 41% (AAPOR Rate #3). We will explore the difference in responses to key behavioral and demographic questions by the number of call attempts, compare the final sample to Census statistics for the target region, and provide information on outcome and cost differences for different maximum call attempt scenarios. This analysis will provide useful information for researchers seeking to evaluate the tradeoff

between increased data collection costs and improved data quality, and will contribute further to understanding the relationship between survey effort and nonresponse bias.

#### **4. The Effectiveness of Advance Letters for RDD Cell Telephone Samples**

Benjamin Skalland, *NORC at the University of Chicago*

Zhen Zhao, *Centers for Disease Control and Prevention*

Jenny Jeyarajah, *Centers for Disease Control and Prevention*

Abera Wouhib, *Centers for Disease Control and Prevention*

In random digit dial (RDD) telephone surveys, advance letters mailed prior to dialing sampled telephone numbers may increase survey response rates (de Leeuw et al. 2007). The ability to mail advance letters to RDD samples relies on the availability of addresses that matched to the sampled telephone numbers. Traditionally, address matching was possible only for landline telephone samples with directory listings, which are not generally available for cell telephone numbers. It is now possible to obtain mailing addresses for a sizeable proportion of cell telephone numbers. Since cell telephone samples are now an increasingly large part of RDD telephone surveys, the use of advance letters mailed prior to dialing cell telephone numbers may result in an increase in response rates similar to those seen for landline telephone numbers. To test this possibility, mailing addresses were obtained for samples of landline and cell telephone numbers in the 2013 National Immunization Survey, a large, national, dual-frame RDD survey sponsored by the Centers for Disease Control and Prevention and fielded by NORC at the University of Chicago. Prior to dialing, advance letters were mailed to half of the cases in the landline and cell telephone samples with available addresses. In this study, we compared address match rates and address accuracy rates between the landline and cell telephone samples and measured the effect of the advance letter on survey response and eligibility rates in the landline and cell telephone samples. We found that while advance letters had a small, positive effect on landline sample response rates, they did not impact cell telephone sample response rates. The lack of effect in the cell telephone sample may be due to a higher rate of inaccurate address matching than in the landline telephone sample. Implications of these results for dual-frame RDD telephone surveys will be discussed.

#### **5. Catch Them If You Can: Tailoring Recruitment Strategies to a Hard-to-Reach Demographic**

Anh Thu Burks, *The Nielsen Company*

Oana Dan, *The Nielsen Company*

The survey research literature has long observed and tried to address the decline in response rates among the general population. Potential respondents have become immune to the “one size fits all” survey requests, which are easy to ignore unless they immediately resonate with respondents’ interests or reach them in their “every day” environment. It is particularly difficult to gain the cooperation of demographic groups like the college-age cohort (18-25 year-olds). An efficient and potentially impactful way to catalyze acceptance rates among this demographic is tailoring the recruitment strategies, motivational messages, and communication channels to their preferences. For example, young adults respond better and quicker to electronic communication by email, text or website, than to written materials. This paper discusses a new strategy to reach and recruit 18-25 year olds, based on feedback from this demographic. Insights gathered from phone surveys and focus groups with college-age students were leveraged to devise a recruitment strategy adapted to the needs and preferences of young adults. We assessed and optimized the content and

vehicle of delivery of the materials used to recruit college students in a long term research panel. The recruitment strategy of incorporating new materials and shifting to electronic media, was deployed in fall of 2013. This paper presents the data used to arrive at the new recruitment strategy, as well as the metrics implemented to assess its effectiveness. We tracked acceptance and refusal rates before and after the implementation of the new strategy, as well as the specific reactions to the new materials deployed. We assess if tailoring the motivational messaging and communication channels to respondents' changing preferences increases the participation of hard-to-reach demographics in research panels, and if these findings point toward the development of more flexible communication channels, easily adaptable to the target audience.

## **6. Tradeoffs in Quality: Examining the Relationship Between Cell Telephone Respondent Location and Data Quality**

Christopher D. Ward, *NORC at the University of Chicago*

Becky Reimer, *NORC at the University of Chicago*

Laurie Elam-Evans, *National Center for Immunization and Respiratory Diseases, CDC*

David Yankey, *National Center for Immunization and Respiratory Diseases, CDC*

Meena Khare, *National Center for Health Statistics, CDC*

Incorporating a cell telephone frame in surveys poses multiple challenges to survey researchers. It is becoming increasingly important to include cell telephone samples as the proportion of cell telephone-only and cell telephone-mainly households continue to increase in the United States. However, cell telephone samples often have lower response rates and higher operational costs than landline telephone samples. This research examined another dimension affected by the inclusion of cell-phone frames: data quality. In particular, we examined whether data quality varied by respondents' sample frame and location (landline telephone-at-home, cell telephone-at-home, and cell telephone-away) at the time of the interview. We used data from all four quarters of the 2012 National Immunization Survey, a national, dual-frame random digit dial survey (analytical subsample:  $n = 15,067$ ). Three types of data were evaluated to examine the relationship between sample frame and data quality: respondents' ability to recall healthcare provider names and contact information, the reporting of vaccinations their child received, and the completeness of reported sociodemographic characteristics. Preliminary results from analyses of variance and tests of proportions showed that data quality may be lower for respondents who were away from home at the time of the interview. This research provides insight into the differential reporting of certain information provided by cell telephone-away respondents and the conditions under which these cell telephone-away respondents may be most likely to provide complete, accurate data. Given the differences in data quality among landline telephone-at-home, cell telephone-at-home, and cell telephone-away respondents, we discuss limitations and implications for the types of telephone survey questions or information that may be problematic for data quality when interviews are administered to cell telephone-away respondents.

## **7. The Impact of Deadlines on the Effectiveness of Incentives**

Morgan E. Haight, *Fors March Group, LLC*

Due to decreasing participation in voluntary surveys among the general and youth populations, research has focused on the use of incentives to encourage survey

participation. While incentives have been shown to increase response rates and minimize bias, do deadlines decrease the effectiveness of contingent incentives? This poster presents the results of an experiment testing the impact of introducing a deadline for incentives in a nationally representative mail survey of youth ages 16 to 24. Participants were sent letters containing an initial \$2 incentive with the opportunity to receive an additional \$5 after the survey was completed. The experiment varied the letters describing this contingent incentive by stipulating a completion deadline to earn the incentive (experiment) with no mention of a deadline (control). Respondents were given approximately 2.5 months to complete the survey and earn the contingent incentive.

Analysis of the experiment's results will be done by comparing outcome rates and survey responses across conditions. Preliminary results showed that response rates were higher for respondents who received letters without the deadline, consistent with existing theory of social exchange in that the introduction of the deadline might increase the perceived costs of responding relative to anticipated rewards. Additionally, we will explore whether the mention of a deadline impacts response timeliness, key metrics for this survey, or survey quality (e.g., item nonresponse). Therefore, survey researchers offering incentives should be thoughtful about how they communicate any contingencies on incentives to participants. Requirements such as deadlines are likely to negatively impact survey responses.

## **8. Why Won't They Respond? Reasons for Respondent Resistance as Recorded by Interviewers in the American Community Survey**

Mary Frances E. Zelenak, *U.S. Census Bureau*

As survey response rates continue to decline, methodologists are trying to determine why respondents are unwilling to participate in their surveys and researching ways to address their concerns. In the American Community Survey, sampled households can respond by one of four sequential modes: answering survey questions online via the Internet, completing and mailing back a paper questionnaire, talking with an interviewer by telephone, or speaking with an interviewer during a face-to-face interview. Among these, the telephone and personal visit modes, which are conducted by interviewers, provide an opportunity to gain some insight into why respondents are reluctant to participate in the American Community Survey. In both modes, the computer-assisted interview includes the survey questions for the respondent as well as some additional questions for the interviewer including whether any resistance was demonstrated by the respondent. This paper will look at the types and frequency of respondent resistance in the American Community Survey recorded by interviewers during a specified period, note how well interviewers were able to overcome the resistance, and provide additional suggestions for ways to reduce respondent resistance.

## **9. Use of Block Group Information to Disproportionately Sample Hard to Reach Demographics**

Lukasz Chmura, *The Nielsen Company*

Scott Bell, *The Nielsen Company*

Tracie Yancey, *The Nielsen Company*

In an effort to improve the quality of address based sampling, The Nielsen Company investigated the use of Census Block Group information to disproportionately sample the following hard to reach demographics: Age of Head of Household (AOH) < 35, Black race, and Hispanic identity.

The above-mentioned groups are less likely to cooperate in a survey when compared to other demographic groups. Consequently, in a random study of the general population, these groups are likely to be under-represented, requiring the use of special treatments, such as differential incentives, and/or post-stratification weighting.

An alternate solution is to oversample those Census Block Groups with higher penetration of AOH<35, Black race, and/or Hispanic identity households. By increasing the probability of selection for households within these Census Block Groups, we can achieve a final reported sample that is more representative of the general population.

This approach was implemented for the Nielsen TV Diary Service in November 2013, and results are presented comparing an address based sample with Census Block Group oversampling vs. one without.

## **10. The Impact of Incentive Levels on Subsequent Respondent Behavior**

Carolyn J. Wilke, *University of Michigan*

Ting Yan, *University of Michigan*

This research examines the impact of varying the level of an offered incentive at wave one of a panel survey on respondent behavior at wave two. It seeks to understand if high levels of incentives encourage good behavior from respondents, or if extraordinarily high incentives simply prompt individuals who would ordinarily not participate (and are thus less inclined to cooperate throughout the survey) to comply with the survey request. A year's worth of survey and call records data from the Survey of Consumer Attitudes allows examination of wave two respondent behavior. Results indicate that the level of incentives respondents receive at wave one affects many aspects of respondent behavior at wave two, including the required number of callbacks, the level of incentive, and the level of item nonresponse. Wave one incentive level also affects respondent attitudes towards the survey, as measured by interviewer observations. This paper offers implications for balancing survey cost with data quality

## **11. Return to Sender: An Analysis of Undeliverable as Addressed Mail Return Rates in the National Children's Study**

Rachel Carpenter, *NORC at the University of Chicago*

One of the primary challenges in any longitudinal study is attrition and the resulting loss of data. Attrition is a serious concern due to the potential for introducing nonresponse bias. One area of attrition studies that has received little attention is predictors of loss to follow-up due to undeliverable as addressed (UAA) mail. For many longitudinal studies, mail is an important way to reach and track participants. As a result, when mail is returned as undeliverable, it becomes more difficult for study investigators to track participants, and may lead to involuntary attrition. In addition, undeliverable mail makes it difficult for study investigators to distinguish between a non-respondent, who receives a survey but fails to respond, and a participant who does not receive a survey due to a change in address.

This paper provides an analysis of predictors of undeliverable as addressed mail for a subset of participants in the longitudinal National Children's Study (NCS), relying on NCS and Census data. The paper reports on findings from several rounds of mailings that were

sent to NCS participants by investigators at NORC at the University of Chicago beginning in March 2013.

This analysis tests whether four key variables of interest were associated with an increase in the likelihood that a mail item would be returned as undeliverable. The four variables were: 1) Percent of the county that is identified by the Census Bureau as rural; 2) Whether the participant moved out of state since being recruited into the Study; 3) Whether a participant had different home and mailing addresses; and 4) Length of time since last completed study visit. This paper discusses the implications of the findings for the NCS, as well as some of the potential limitations of the data and ideas for future research.

## **12. Evaluating the Effectiveness of Contact Strategies to Improve Response to the 2012 Economic Census**

Jennifer Beck, *U.S. Census Bureau*

In addition to a population census, the U.S. Census Bureau conducts an Economic Census every five years in order to measure the health of businesses and the U.S. economy. The Economic Census is a large data-collection operation that includes more than 500 questionnaire versions for nearly four million businesses. For the 2012 Economic Census, the Census Bureau conducted a series of experiments embedded within data collection in an effort to evaluate alternative methodologies for improving response and saving costs. The experiments focused on different contact strategies aimed at improving response rates and the timeliness of response and encouraging Web response. These strategies included using advance letters, due-date reminder phone calls, and letter-only follow-ups without a replacement questionnaire. The preliminary results reveal that sending an advance letter to targeted industries with lower response propensities did not have a significant effect on response rates. Automated phone calls to remind respondents of the upcoming due date also did not have a large effect on the overall response rates. However, following up with non-responding businesses using only a letter that encouraged Web response and without including a replacement questionnaire increased the proportion of Web response. This effect was larger for new businesses that previously had never been in sample for the Econ Census or any other Census Bureau survey. This paper will discuss these results in detail and their implications for incorporating adaptive design strategies into data collection for future Economic Censuses.

## **13. Are Conditional Monetary Incentives with an On Demand Pay-Out Option a Cost-Effective Measure to Reduce Panel Attrition? Findings From a Field Experiment**

Ines Schaurer, *GESIS - Leibniz Institute for the Social Sciences*

Michael Bosnjak, *GESIS - Leibniz Institute for the Social Sciences*

Conditional monetary incentives are often seen as a measure to reduce attrition in panel surveys. However, they constitute a substantial share of survey costs. To reduce these costs, most panel providers use on demand pay-out options, hoping that not all panelists redeem their incentives. We seek to explain panel attrition/retention as a function of incentive amounts, controlled for actual incentive redemption. In doing so, we analyze the data from two offline-recruited probability-based panels conducted by GESIS.

The first panel is an online panel of Internet users surveyed monthly for 8 months starting in February 2011. The amount of 0, 2, 5 and 10 EUR conditional incentives was varied

experimentally. Incentives were paid out on demand as cash, voucher, or donation for charity.

The second data come from the GESIS Panel, a mixed-mode (online and mail) panel of the general population in Germany. The recruitment started in June 2013, respondents are surveyed bimonthly. The panel implements prepaid initial incentives and conditional incentives after completion of the initial questionnaire.

For the first panel there is an overall positive effect of incentives on panel retention. However, only about half of potential incentive recipients use the redemption option. The probability of payment increases with every completed wave (Odds Ratio=2.47;  $p<0.001$ ) and incentive amount ( $\text{Chi}^2(3)=41.99$ ,  $p<0.001$ ). Still among those who completed all 8 waves, one third of panelists never initiate incentive payment.

For GESIS Panel, only a minor part of respondents to both the first and the second waves wish to redeem the incentive. With more waves fielded, it will be possible to say at what rate respondents want their incentives paid out.

Overall, the findings indicate that conditional monetary incentives on demand are both fair to respondents and cost-effective for panel management. We demonstrate ways to implement such incentive schemes, and sketch future research directions.

#### **14. Best Practices in Recruiting Parents of Children in Child Care Programs for Research Study Participation**

Valerie Atkinson, *Westat*

Kwang Kim, *Westat*

Christine Nord, *Westat*

Collecting data from parents of children in child care programs can be challenging, since the amount of time they have to talk to a recruiter at the child care center is often limited and busy with competing demands. This poster shares lessons learned from the Family-Provider Relationship Quality (FPRQ) study in recruiting parents.

The purpose of the FPRQ study is to develop measures that capture aspects of the family-provider relationship found to be associated with positive child and family outcomes. These measures are being developed for use in center-based and family child care programs and to be culturally appropriate for diverse populations, including lower- and higher-income families, ethnically/racially diverse providers and families, and non-English-speaking families and providers. Three surveys are in development: director, provider/teacher, and parent questionnaires.

A pilot test of the FPRQ study was conducted in early 2013. The original parent data collection procedures for the pilot test involved leaving informational flyers at the programs with a phone number for parents to call to participate. This passive approach to parent recruitment was largely unsuccessful, and to ensure adequate parent participation in the study, recruiters later modified these procedures to more actively engage with parents. The revised procedures seemed to increase parent cooperation, and suggest that more promising recruitment procedures might be considered for future research.



In early 2014, we will conduct a large-scale field test. Data collection procedures for this field test will use direct approaches to parent recruitment, including onsite recruitment of parents through face-to-face contact.

Our goal is to share our experiences and results from the field test (completion April 2014) to help other researchers with similar data collection challenges by demonstrating best practices for recruiting parents and similar populations for research studies.

## **15. Accompanied Face-to-Face Interviewing as Data Validation: Does Accompanied Interviewing Affect Data Quality in Countries with Underdeveloped Democracies?**

Jenna Levy, *Gallup, Inc.*

Dato Tsabutashvili, *Gallup, Inc.*

Neli Esipova, *Gallup, Inc.*

Previous studies have long provided evidence that survey error increases as the number of people present during the interview increases. This error is derived from biases in both measurement (social desirability or confidentiality concerns) and item nonresponse, particularly among sensitive questions. While there are numerous studies that focus on the effects of measurement and item nonresponse error, there is a gap in the literature regarding effects on data quality in countries with underdeveloped democracies. This research investigates the implications of supervisor-accompanied interviews on data quality, particularly regarding items about political, social, and moral attitudes in the CIS (Commonwealth of Independent States) region via Gallup's 2011-2013 World Poll data.

The Gallup World Poll conducts surveys in 164 countries around the world. To attain high levels of coverage, Gallup collects data using face-to-face methodology in many countries where telephone coverage is not sufficient. As part of the data collection process, Gallup employs a variety of validation techniques including accompanied interviewing.

Accompanied interviews are defined as a supervisor attending the interview to ensure the interviewer is properly administering the survey. This research seeks to understand the ways in which accompanied interviewing might affect quality of different types of survey questions, and also make a recommendation for whether to use accompanied interviewing as a method for validation in the future.

This study will conduct analysis using country-level data from CIS countries with at least 10% of the sample with the accompanied interview treatment: Russia, Ukraine, Uzbekistan, Tajikistan, Turkmenistan, and Belarus. The research will use individual items with varying degrees of cultural sensitivity in each country to address whether accompanied interviewing increases socially desirable reporting and item nonresponse, or oppositely, increases interviewer performance.

## **16. Making Sense of Paradata: Challenges Faced and Lessons Learned**

Adam Eck, *University of Nebraska-Lincoln*

Leonard Stuart, *University of Nebraska-Lincoln*

Gregory Atkin, *University of Nebraska-Lincoln*

Leen-Kiat Soh, *University of Nebraska-Lincoln*

Allan L. McCutcheon, *University of Nebraska-Lincoln*

Robert Belli, *University of Nebraska-Lincoln*

Paradata collection has become increasingly popular as a means to provide additional information and context to augment traditional data collection in survey methodology research, especially in computer and Internet-based systems (Kreuter 2013, Callegaro 2013). Paradata enables researchers to analyze aspects of the survey or interview process that relate to data collection (e.g., respondent timeliness, question navigation order), and even recreate the process a posteriori based on log files tracking respondent and interviewer behaviors. However, understanding and making use of paradata can be quite challenging. First, paradata commonly represents information at a much lower level of abstraction than the survey or interview data being collected (e.g., mouse clicks, keyboard presses, software actions), requiring careful mapping between paradata and higher order data. Second, due to its fine granularity, multiple instances of paradata must often be considered in combination (rather than individually) to extract useful meaning (e.g., response times between two actions). Third, as the size of survey data grows, managing and linking each type of information becomes more challenging, especially across different levels of granularity. Finally, paradata representation can vary wildly between surveys and computer systems, making it difficult to develop a standard set of tools and techniques for working with paradata. In this presentation, we describe the challenges faced and lessons learned from working with different types of paradata across two different surveys: the web-based portion of the Gallup survey panel and the interview-based American Time Use Survey (ATUS). We propose methods for (1) organizing and relating paradata to survey data (using a relational database model for scalability), (2) extracting useful information from paradata (recreating the state of surveys and interviews), and (3) guidelines for paradata generation to facilitate usage during analysis. We also describe ongoing results and benefits of employing these methods in our analysis of these two surveys.

## **17. Call Back Later: The Association of Recruitment Contact and Error in the American Time Use Survey**

Austin Countryman, *University of Nebraska-Lincoln*

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

Caitlin E. Deal, *University of Nebraska-Lincoln*

Robert F. Belli, *University of Nebraska-Lincoln*

Respondents to time diary surveys are often poor at remembering what they did even a few days earlier (Bolger et al., 2003). For this reason, time diary surveys generally ask only about more recent events, for example, those that occurred the previous day. As with any survey, respondents vary greatly in both their motivation to complete the survey, and the number of contact attempts required to achieve completion. In fact, these factors are likely related, as late responders may tend to be less motivated than their early responding counterparts (Donald, 1960; Paganini-Hill et al., 1993). Furthermore, a lack of motivation is associated with failures in the cognitive processing of survey questions (Krosnick, 1991). The retrieval phase of the response process, for instance, may fail if the respondent is not sufficiently motivated, and thus more recall errors may occur (Tourangeau et al., 2000).

I hypothesize that late responders, compared to early responders, will produce more recall errors among respondents to a time diary survey, signaling lower motivation. The cognitive shortcuts used to bypass retrieval will produce higher levels of measurement error, and lower quality data. Error will be assessed from the quantity of respondents' recall failures of travel and activity type in the American Time Use Survey (ATUS), a nationally representative time diary conducted on a 2010 sample.

As expected, preliminary analyses indicate that there is a difference between late and early responders; specifically, compared to early responders, late responders are significantly more likely to fail to report travel for each additional week during the 8-week response period in which a call attempt was made. These results suggest that late responders may be less motivated than early responders, and therefore more likely to employ cognitive shortcuts. Further analyses will examine whether this relationship extends to other ATUS errors as well.

## **18. An Experimental Comparison of a Web, Telephone and Paper Survey with Older Adults**

Chan Zhang, *University of Michigan*

Lindsay H. Ryan, *University of Michigan*

Mary Beth Ofstedal, *University of Michigan*

Richard D. Gonzalez, *University of Michigan*

Jacqui Smith, *University of Michigan*

Traditionally, Web surveys have been considered to be less feasible for older compared with younger adults. But this picture is changing. As of 2012, it is estimated that about 80% of American adults aged 50-64 and half of those over 65 use the Internet. However, it is still unclear whether online response quality for older adults differs from that of traditional survey modes. Prior studies comparing Web and other modes have focused on general populations. We speculate that two age-related factors – cognitive decline and relatively less experience with Web surveys – may lead to more pronounced mode effects among older adults. We conduct a randomized experiment to compare Web with Telephone and Paper modes for surveying older adults (data collection is ongoing). 240 locally recruited participants (age over 50 and users of email) are asked to complete two surveys separated by one month about their experiences yesterday. Participants are randomly assigned to one of the four paired mode conditions (Web first – Paper second, Web first – Telephone second, Telephone first – Web second, Paper first – Web second). This design allows analyses of between-group and within-person mode differences. We focus on two response quality indicators: differentiation in ratings with the same scales and item missing. We hypothesize less differentiated ratings on telephone (because of the need to memorize the scale) compared to those in Web and Paper questionnaires. Further, we speculate this effect may be even larger among those over 65 due to normative age-related memory decline. We expect item missing to be higher in the two self-administered modes than in telephone surveys, and potentially highest in Web if respondents accidentally miss questions as they navigate through the questionnaire. This could especially be the case for the oldest adults who tend to be less familiar with Web surveys.

## **19. Continuing Investigation of Attitudes Towards the Use of Administrative Records**

Ryan King, *U.S. Census Bureau*

Jennifer Hunter Childs, *U.S. Census Bureau*

Monica Wroblewski, *U.S. Census Bureau*

In reaction to declining response rates and increased operational costs, the Federal Statistical System is carefully examining the possibility of using administrative records to supplement current survey practices. To do this, we need to understand what the public's reaction may be and what concerns the public may have if this is undertaken. An interagency team developed a series of questions that explore various situations and perceived benefits for administrative record use, which are asked on an ongoing nightly telephone survey. Past research using this tool suggested that while respondents were somewhat split on the use of administrative records by a number of Federal agencies, but when we informed respondents that use of the record could save the government money or provide a societal good, respondents were more likely to be in favor of their use.

This paper contains the results of a follow-up to this research. In this implementation, we focused only on a single agency requesting the use of an administrative record, but we varied whom the source of the record was and what type of information was being requested so we could isolate the effects of each of these. We also varied the perceived benefit of the message in which we inform the respondent about helping to save the government money (using an overall or per unit savings) and providing a societal good (at the local or national level). The paper will investigate whether there are differences in opinion towards administrative records by source of the record or type of information being requested. In addition, we show how respondents of different demographic groups and of different mindsets may have different attitudes towards the use of administrative records depending on how the use is framed.

## **20. Qualitative Results that Inform Quantitative Paradata**

Elizabeth M. Hoeffel, *U.S. Census Bureau*

There is a great deal of both qualitative and quantitative research done in the field of survey methodology, but often they are not used in a complementary manner. This research combines qualitative results from post-collection respondent debriefings for the 2012 Economic Census Respondent Debriefings with examination of paradata from its web questionnaire. Given the volume and complexity of paradata generated by web survey instruments, being able to focus analysis may guide more efficient use of time and resources. We used findings from debriefings with single establishment businesses to help us identify respondents' issues with the features and performance of the instrument, specify research questions, and prioritize and focus analyses of associated paradata. In the debriefings, among other things, respondents mentioned they did not know the difference between warnings and errors, had issues with submitting the data, and when asked about future mailings, stated that they would like to be able to record information by writing on hardcopy paper forms before entering the data into the online instrument. In this paper, we examine the paradata for evidence and extensiveness of these respondent-reported concerns and behaviors. Using quantitative data to augment qualitative findings could provide a stronger evidence of issues that need to be addressed to improve data quality and reduce respondent burden.

## **21. Experimental Comparison of Different Incentives for Recruiting for Qualitative Studies**

Patricia LeBaron, *RTI International*

Bridget Kelly, *RTI International*

Lauren McCormack, *RTI International*

Marjorie Margolis, *RTI International*

Dhuly Chowdhury, *RTI International*

Identifying the best methods to recruit qualitative research participants requires a number of design considerations. Studies on survey response rates suggest cash incentives perform better than tokens or lotteries, but results of comparisons of these noncash incentives to no incentive are mixed (Boyle & Heyworth, 2012). To assess the impact of different incentive amounts on willingness to participate in qualitative research, we conducted an experimentally designed study of the impact of different incentives on respondents' willingness to participate in a hypothetical qualitative research effort.

We recruited 4,136 members of GfK Custom Research's Internet panel in July of 2013. Respondents viewed a flyer recruiting participants for an hour-long in-person interview about health topics. Respondents were randomized to one of five incentive conditions, including no mention of incentive, the offer of entry into a lottery, and a range of cash incentives up to \$75. We then measured willingness to participate in the research.

Results show that the odds that someone would be willing to take part in the research generally increased with the incentive amount offered but leveled off at \$50. Those who viewed the version of the flyer offering the nonmonetary incentive were no more likely to be willing to participate than to those who received the \$0 incentive ad. The odds of being willing to participate in the interview for the \$25, \$50, and \$75 incentive amounts were all significantly higher compared to the \$0 group.

Limitations include that the sample consists of people who have already agreed to be members of an Internet panel and are likely to be more altruistic than the general population. Also, this study reflects a hypothetical situation; people may act differently when confronted with an actual invitation to participate in research.

These results provide guidance on optimal incentive levels for recruiting qualitative research participants.

## **22. Effect of Using \$1 Incentive on Response Rate, Response Pattern and Sample Composition – Evidence from an ABS Mail Survey of Fishing Effort**

Sujata Pal, *Abt SRBI*

Courtney Kennedy, *Abt SRBI*

Alex Shapiro, *Abt SRBI*

Marci Schalk, *Abt SRBI*

Rob Andrews, *NOAA Fisheries*

Given challenges with RDD data collection and costs, many researchers are exploring mail surveys with ABS samples drawn from the U.S.P.S. Computerized Delivery Sequence File. Researchers are beginning to shed light on how to optimize the recruitment protocol for such surveys, particularly those that screen for a specific subpopulation (e.g., Brick et al.

2011; 2012). One aspect of ABS recruitment for screening surveys that is being researched is whether to offer an incentive and, if so, what amount.

In this study, we present results of introducing pre-paid \$1 incentive in the 2012 NOAA Dual Frame Mixed Mode Survey of Fishing Effort. The survey was conducted over six waves in Florida, Georgia, North Carolina and South Carolina. In each wave, a mail screening survey was conducted using a random ABS sample design which was used to recruit respondents eligible for a subsequent mixed-mode fishing effort survey. Although no incentive was used for the first three waves, a \$1 pre-paid cash incentive was introduced in Wave 4 of the survey and was continued through the rest of the study.

We compare the effects of introducing the incentive in terms of response rates, response patterns, response time and composition of the sample. Results suggest that inclusion of a \$1 cash incentive not only led to a large increase in response rate, but also significantly reduced mean response time. In a two stage data collection design where a screener survey is used to recruit respondents for a subsequent survey, a reduced response time can improve sample quality as well as reduce data collection costs.

### **23. Identifying Doubled-Up Households Using Survey Data**

Kate E. Bachtell, *NORC at the University of Chicago*

Catherine Haggerty, *NORC at the University of Chicago*

The U.S. Census Bureau has reported a significant increase in the number of doubled-up households following the 2007 economic recession, including a 2% growth in the proportion of young adults ages 24 to 35 living in their parents' homes between 2007 and 2009. These households are defined as those including at least one non-student adult who is not the householder or the householder's partner (DeNavas-Walt et al 2011). In 2011 18.3% of U.S. households were doubled up - an increase of 1.3% since the height of the housing market boom in 2007. These data have inspired many studies examining doubling up as a strategy for making ends meet during times of financial hardship. Analysts are often challenged to identify doubled-up households using household roster data, without the benefit of contextual information about life cycle events (marriage, new births, etc.) and despite the temporary nature of many doubled-up housing arrangements. Data limitations may confound efforts to create a measurement of doubling up that captures substitutions of individual household members and other complexities. In this paper we present findings from a small meta-analysis of techniques used to isolate doubled-up households with survey information. We describe our experiences working retroactively with longitudinal data from the Survey of Consumer Finances and Making Connections Survey and offer recommendations for best practices at both the data collection and analysis stage.

### **24. Exploring a Method to Evaluate Survey Response Scales**

Rene Bautista, *NORC at the University of Chicago*

Lisa Lee, *NORC at the University of Chicago*

This study presents results from a qualitative evaluation of two alternate survey scales. NORC at the University of Chicago, through a subcontract with MITRE for the Office of the National Coordinator for Health Information Technology, conducted a survey of consumer attitudes toward health information technology. In the Year 1 survey, respondents were presented with a four-point concern scale for questions on privacy and security of electronic health records (very concerned—somewhat concerned—not very concerned—not

concerned at all). Across these questions, 25%-50% of respondents expressed the highest level of concern (very concerned). It was not clear whether this was a ceiling effect, in which respondent concern was truly at the highest level, or whether this reflected the need for an additional option between the categories of very and somewhat concerned. We conducted 25 cognitive interviews, 15 in English and 10 in Spanish, to test an alternative five-point scale that included the option of fairly concerned (very concerned—fairly concerned—somewhat concerned—not very concerned—not concerned at all). Respondents were first presented the questions on concerns about privacy and security issues with the four-point scale. They were then asked whether they would have chosen fairly if this option had been available. Respondents were also asked to visually represent the distance between options on the four- and five-point scales by marking on a line where they would place each of the response categories. We found that the distances between response options were not always equidistant. Further, a number of respondents chose to switch the ordering of fairly and somewhat because they felt that somewhat was a stronger term than fairly. Further, the cognitive interviews highlighted some issues that were unique to the translation of the scale into Spanish.

## **25. I Think I Know You: A Quasi-Experimental Study Showing the Effectiveness of Area Code Matching for Telephone Surveys**

Heather Knappen, *Metrix Matrix Inc.*

This poster will present a quasi-experimental, pre-post designed study on the effect of area code matching on telephone survey response rates within the commercial sector. The prevalence of caller ID technology used for personal call screening often impedes survey response rates among respondents who are unlikely to answer calls from unfamiliar telephone numbers. This study asks if area code matching helps overcome this issue by increasing response rates.

In 2012, Metrix Matrix instituted a method for smart dialing within its in-house call center. Smart dialing relies on purchasing phone numbers within targeted area codes. These numbers, with the desired area codes, are displayed on outgoing calls made to the same, or geographically adjacent, area codes. The data presented in this poster will span a period of three years, from 2011-2013, in an attempt to capture the effect on contact and cooperation rates before and after the implementation of smart dialing. A randomly generated sample of 10,000 phone calls from each of these three years will be analyzed by year, cell phone v. landline respondents, and various industry sectors that include utilities, business, government and healthcare.

This study is one of the first that tracks contact and cooperation rates as well as the effectiveness of area code matching for telephone surveys. As the AAPOR community continues to evaluate the role of cell phones in survey research, the use of smart dialing may be a promising tool to encourage participation among cell phone respondents.

## **26. A Review of Households Experiencing an Interruption in Telephone Service**

Randal ZuWallack, *Abt SRBI*

As the telecommunications landscape continues to transform, our research methods evolve to keep pace. More people have cell phones than landlines; the cell-only population is growing; the landline-only population is shrinking. But one constant throughout the past decade is the percentage of nontelephone households, steady around two percent.

Nearly two decades ago, Keeter (1995) introduced “transient” telephone households, or those who periodically experience interruptions in telephone service, and found that they were similar to nontelephone households (then about five percent) along a number of socio-demographic dimensions. His findings led to weighting adjustments to reduce the risk of coverage bias due to nontelephone households. These adjustments leverage sampled households who have experienced interruptions in service to represent those currently without service.

The question is, with the current state of telecommunications, do these adjustments still help? According to the National Health Interview Survey (NHIS), the percentage of landline households experiencing an interruption in landline telephone service has averaged 3.5 percent since 2007 (one week or more; past 12 months). But of those that did, nearly 80 percent reported that they had a working cell phone during that time. Our research reviews interruptions in telephone service from the NHIS and a dual-frame telephone sample from the National Highway Transportation Safety Administration’s National Survey of Bicyclist and Pedestrian Attitudes and Behaviors. In this survey, landline and cell respondents were asked to report whether they were without any telephone service (landline or cell phone) for one week or more in the past 12 months. Over five percent (n=394) report they were, many of whom were cell-only (n=152). We compare the socio-demographic characteristics of these respondents and evaluate them as a proxy for non telephone households.

## **27. Tweets of Kindness: Understanding the Relationship Between Social Media Expressions of Benevolence and the Translation to Real World Behaviors**

Reagan B. Barbee, *East Tennessee State University*

Kelly N. Foster, *East Tennessee State University*

Does expressing support for something through social media actually reduce the likelihood that a person will offer support off-line? If a person offers prayers for victims of a national disaster do they feel as though they have satisfied their social obligation to do something thereby actually reducing the amount of tangible support offered? Social media allows people to share all aspects of their daily lives and things that are important to them. However, the extent to which the online persona matches offline realities is unknown. In May of 2013, the state of Oklahoma experienced one of the worst tornados in recorded history that resulted in \$2 billion dollars in damages, 25 deaths, 377 injured and countless other strains on the social and economic landscape of the state. In response to this tragedy, many people took to social media to express their sympathy and concern for the citizens of Oklahoma. A popular hashtag that emerged was #PrayersforOklahoma in which all manner of private citizens and public figures tweeted their support and concern. Comedian Ricky Gervais took issue with the social media support stating he felt individuals should provide monetary or physical support for victims instead of only sending prayers through Twitter. This “twitter war” between #prayforoklahoma and #actuallydosomethingforoklahoma lends itself to an interesting research project to examine if expression of benevolence through social media actually manifests itself into tangible acts of kindness off line. For this project, individuals who tweeted #PrayersforOklahoma and #actuallydosomethingforoklahoma are contacted via Twitter and asked to complete a web-based instrument to determine what, if any, off-line support they may have offered to victims of the Oklahoma tornado. This research seeks to add to the body of literature that examines how a person’s online activities may or may not translate into off-line activities.



## **28. Recruiting the Recruiters: Does Invoking Prior Commitment Increase Teacher Participation and Student Responses in a School-Based Survey?**

Mary E. Losch, *UNI Center for Social & Behavioral Research*

Neal Pollock, *UNI Center for Social & Behavioral Research*

Conducting school-based surveys poses numerous challenges. Unlike standard survey designs, school-based surveys of students must often be filtered through classroom teachers. This additional step can limit success in reaching students (and parents who must provide participation consent). In the context of an evaluation of an externship program for teachers (n=56), two recruitment designs were tested. In one, the invitational messages to the teachers referenced the previous commitment that teachers made to assist with the evaluation. In the other group, no such reference was included. Analyses of teacher participation and student response rate differences will be compared between the two groups to determine the value of references to previous commitment to participate in the evaluation and facilitate the student survey.

## **29. Evaluation of Strategies for Targeting Local Area Cell Phone Samples**

Dean Williams, *Abt SRBI*

Seth Brohinsky, *Abt SRBI*

Marci Schalk, *Abt SRBI*

Randal ZuWallack, *Abt SRBI*

There are two main strategies for targeting local areas with cell phone random digit dialed samples, 1) based on the location of the rate (billing) center the phone number was originally assigned, and 2) the zip code of the current billing address. When determining the targeting strategy, one must balance geographic accuracy with population coverage. Cell phone frames are often determined by selecting a sample from rate centers that fall within the geographic area. This frame definition will result in undercoverage if a high proportion of the population is served by rate centers outside the geography. However, including these outside rate centers outside may be cost prohibitive due to lower geographic accuracy. Similarly, excluding zip code match-outs will increase efficiency, but could have coverage implications.

This presentation examines the accuracy of rate center and zip matching for targeting four cities in the middle of major metropolitan areas: Boston, MA; Los Angeles, CA, Philadelphia, PA; and Washington, DC. First, we present rate center accuracy for all four cities and zip matching accuracy for LA and Boston by comparing with self-reported data. In DC and Boston, we examine the coverage of the city population from rate centers outside the city limits. Finally, we discuss a method for using the zip matching results to estimate rate center incidence and coverage to evaluate the sampling frame.

## **30. Does the Graphic on an Envelope Increase or Decrease Response Rates?: An Experiment with an Address-Based Sampling (ABS) Mail Screener Form for the California Health Interview Survey (CHIS)**

Royce Park, *UCLA Center for Health Policy Research*

Matt Jans, *UCLA Center for Health Policy Research*

David Grant, *UCLA Center for Health Policy Research*

Sherman Edwards, *Westat*

John Rauch, *Westat*

Survey methodologists are always looking for ways to make the material they mail to respondents more likely to be opened, completed, and returned. Traditional advice is to avoid making the delivery envelope look too much like marketing or fundraising for fear that respondents will consider it junk mail, but there have been few experimental tests of this hypothesis. On the other hand, distinctive graphics and messages may be necessary to distinguish survey materials from other types of information and requests people receive in the mail. In this small experiment, we manipulated the delivery envelope and presence of a packet insert in the 2nd mailing of an address-based sample (ABS) screener form asking for basic household information and a telephone number for a subsequent telephone survey. The experimental envelope differed from the standard envelope only in the addition of the survey sponsor's logo (emphasizing public health). Addresses were randomized to receive either the standard (blank) envelope or the experimental envelope (with logo). The experiment was conducted on a random sample of addresses from the USPS Delivery Sequence File (DSF) in two California communities, and used the California Health Interview Survey (CHIS) telephone interview, which includes questions about health behaviors, health status, and insurance status. The screener forms and mailing materials were printed in English and Spanish because the communities sampled both had relatively high rates of Spanish-speakers. Results suggest that traditional advice about avoiding logos in mail survey design may be sound, but it is not clear whether the advantage of the logo-less envelope is systematic. We explore differences between the two communities, and between English and Spanish respondents. Implications for other mail survey designs, particularly those using ABS designs in hard-to-survey communities will be discussed.

### **31. Impact of Presence of Others on Face-to-Face Survey Responses**

Anita Pugliese, *Gallup, Inc.*

Neli Esipova, *Gallup, Inc.*

Dato Tsabutashvili, *Gallup, Inc.*

In international face to face surveys it is relatively common for additional people to be present during the survey. Across 106 countries surveyed in 2012, Gallup's World Poll finds that overall half of all face-to-face surveys were conducted with others persons present (adult or children). The presence of others may be a cause of increased measurement error due to social desirability. This paper examines the impact of the presence of others on face-to-face survey responses, including self-reporting of personal economics, personal health, ratings of the country's political situation and local community, acceptance of diversity, etc. The effect of four conditions are examined: no others present, other adult(s) present, children present, both adults and children present. Logistic regression technique is used to estimate the effect of these conditions on each survey item, controlling for respondent age, education, family size, urban/rural, country. Results may differ by country, and may be partly explained by cultural differences in social desirability. We collected additional information in Commonwealth of Independent States (CIS) countries which enable us to conduct more detailed analysis on the effect of gender of others present. Interviews were categorized into three groups: all female, all male, or mixed (male and female). We will discuss the impact of these findings on strategies to minimize the effect of others presence in face to face surveys.

### **32. Parent Survey Response and Completion Rates with Differing Incentive Payments: A Cohort Study from the Universal Preschool Childhood Outcomes Study, Phases 6 and 7 (UPCOS)**

Rebecca Mason, *Mathematica Policy Research*  
Susan Sprachman, *Mathematica Policy Research*

Since 2007, the Universal Preschool Childhood Outcomes Study (UPCOS) has examined the feasibility, reliability, and validity of child development measures in the large, culturally and linguistically diverse population of children in Los Angeles County. This study informs the performance-based contract between Los Angeles Universal Preschool (LAUP) and First 5 Los Angeles by examining the progress of a representative sample of children served by LAUP programs. In Phases 6 and 7, we administered a short Parent Survey that was distributed with the consent forms. This survey gathers information about participants' backgrounds, household routines and characteristics, and risk factors. To incentivize parents to return the surveys, we raffled off two \$25 gift certificates per classroom in Phase 6. In Phase 7, we offered parents of each selected student a \$10 prepaid gift card. This poster will compare response rates across the two rounds.

A larger proportion of parent surveys was returned in Phase 7 than in Phase 6. In Phase 6, we collected 588 surveys from a total sample of 707 students for a return rate of 83%. In Phase 7, we collected 654 surveys from a total sample of 716 students for a return rate of 91%. This poster further compares the completeness of the survey items over both rounds.

### **33. Recruiting a Truly Representative Sample: A Review of Processes Employed During the Universal Preschool Childhood Outcomes Study, Phase 7 (UPCOS)**

Rebecca Mason, *Mathematica Policy Research*  
Susan Sprachman, *Mathematica Policy Research*

Since 2007, the Universal Preschool Childhood Outcomes Study (UPCOS) has informed the performance-based contract between Los Angeles Universal Preschool (LAUP) and First 5 Los Angeles by examining the progress of children served by LAUP programs. For Phase 7 of the study, we have continued to employ a two-tiered approach in recruiting 40 preschool programs: 1) Following an advance letter, the Site Coordinator calls the program to gain their cooperation. 2) If this approach proves unsuccessful, a Field Enrollment Specialist visits the program in person to enroll families in the study. Sometimes, Coordinators are unable to recruit a site over the phone and an in-person visit becomes necessary. Programs usually fall into one of three categories during the recruitment process: 1) Forthcoming sites: relatively available by phone, responsive, and forthcoming with their intentions to participate during the telephone recruitment process. 2) Hard to contact sites: not readily available by phone; often, voicemail is full and they are not responsive to email. 3) Avoidant sites: not readily available by phone and inclined to refuse participation. These sites may not keep appointments or take calls. In some cases they are initially agreeable, but stall in the enrollment process.

In previous rounds of the study, a handful of programs unavailable over the phone were easily recruited by field staff in-person. While this approach generally proved successful when employed, there was neither a protocol indicating the appropriate timing for in-person visits nor a formal tracking system for follow-up. In Phase 7, we have systematically employed a variety of recruitment techniques, including sending field staff to recruit sites in-person. This poster will provide an overview of the effort, methods, and processes employed

for contacting, triaging, and recruiting a truly representative sample of preschools to participate in the study of preschool outcomes.

### **34. Innovative Retention Methods in Panel Research Phase II: Can Use of SmartPhone QR Codes Improve Long-term Panel Participation?**

Andrew Dyer, *ICF International, Inc.*

James Dayton, *ICF International, Inc.*

Minimizing participant attrition is vital to the success of longitudinal panel research. One such example of a longitudinal panel study conducted by ICF is the National Recreational Boating Survey (NRBS), sponsored by the U.S. Coast Guard to ensure that the public has safe, secure, and enjoyable recreational boating experiences. The project features several components, including a “Trip Panel”. The Trip Panel is designed to measure recreational boating exposure. This panel was developed via a dual-frame, dual-mode (Random Digit Dial telephone and mail) design. Respondent contact information includes e-mail address, mailing address, and telephone number.

This presentation will explore ICF researchers’ continuing pursuit to improve panel retention through non-traditional methods, including text messages and non-voice SmartPhone features. Phase II of our pilot involved testing a QR code to permit and encourage panelist communication of contact information changes and to express interest in continued panel participation. Active panelists were randomly assigned to a group to either communicate contact information updates via postage paid postcard or web survey link (control group), or via postage paid postcard, web survey link, or QR code (experimental group).

ICF researchers will analyze the differences in response rates between the control and treatment groups. We will also monitor panelist continuing panel participation rates.

### **35. Asking Questions About Others to Improve Proxy Responding**

Erica Yu, *Bureau of Labor Statistics*

Proxy reporting on behalf of others may be inaccurate when the proxy respondent does not have the required knowledge or must recall the information from memory. It has been suggested that when such reports are based on memory, they are often based on general impressions about the person rather than retrieval of the specific events or details (Menon, 1995). In the context of the Consumer Expenditure (CE) survey, where respondents often report expenses on behalf of other household members, proxy recall is considered to be poor and a low estimate of true expenditures. Previous research by Book and Edgar (2012) investigated whether additional questions could be used to improve proxy respondent recall of others’ expenditures and increase reporting. In Book and Edgar (2012) several questions about household members’ activities (e.g., travel, hobbies) were tested in cognitive interviewing and were found to have successfully cued additional expenditures. The current study builds on Book and Edgar (2012) by standardizing question scripting for use in production settings. Initial cognitive interviewing using standardized question scripting failed to provoke additional expenditure reports. After modifying the scripts to include more detail from the participant about the household member (e.g., referring to John’s hobby of playing tennis rather than referring to John’s hobby), the protocols successfully cued additional expenditure reports. The results from the two stages of cognitive interviewing are discussed in the context of the results from Book and Edgar (2012). The results suggest that providing

detailed cues at the time of recall is effective in prompting respondents to recall additional information about others.

### **36. Potential Indicators of Availability When Using a Three-Call-Attempt Rule for Telephone-Based Contact**

Michael Zeddies, *Institute for Social Research*

Several previous studies have examined the effects of a three-call-attempt threshold for telephone mode contacts with respondents. Recently, the Surveys of Consumers monthly survey has shifted to a three-call-attempt protocol for determining noncontact status. This study utilizes the data from the June 2013 SCA (which had not yet implemented this protocol) and examines the relationship of key variables with availability by the third call attempt. In particular, the June 2013 SCA featured a rider that measured the physical mobility of respondents on several objective measures. Although there was an isolated relationship of respondent mobility with availability by the third call, this relationship disappeared in a more complex model that controlled for age. It is also interesting that past SCA late contacts will in the future become non-contacts. Directions for further study, and potential implications for future telephone contact efforts, including future months of the SCA under its new protocols, are explored.

### **37. An Experimental Test of Prepaid Incentives and Type of Mailer to Increase Telephone Survey Response**

Rebecca DiGiuseppe, *Mathematica Policy Research*

Julita Milliner-Waddell, *Mathematica Policy Research*

Mathematica Policy Research is conducting a study for the U.S. Department of Labor (DOL) that will evaluate the impact of a health benefits subsidy under the Consolidated Omnibus Budget Reconciliation Act (COBRA) that was provided by the American Recovery and Reinvestment Act of 2009. We aim to complete 3,700 CATI interviews with unemployment insurance claimants who were eligible for COBRA when they lost their job.

We send all sample members an advance letter with two options to determine their eligibility for the study: (1) call an IVR system to complete a brief screening survey or (2) call the Mathematica survey operations center directly. To encourage use of the less-costly IVR option, sample members are offered \$40 if they are determined eligible through the IVR and complete the full interview within 30 days of the advance mailing. Eligible sample members who complete the screening survey with an interviewer or call outside the 30-day window are offered \$30.

We developed three experiments using different incentive and mailing strategies to determine the most effective way to increase response. For the first experiment, we sent a letter to a subset of our oldest sample members, who were outside the \$40 incentive window. The letter offered sample members \$40 and was sent in a Priority Mail envelope.

For the second and third experiments, we sent a \$5 prepaid incentive, with a letter promising an additional \$35 for completing the survey. We sent the letters for one experiment group using Priority Mail, and letters for the other group in business-size DOL envelopes.

These experiments are currently underway. This paper will discuss the results and make recommendations for future survey research with similar populations.

### **38. Randomization in Computer Assisted Interviews**

Kien Le, *Social and Economic Survey Research Institute, Qatar University*

Abdoulaye Diop, *Social and Economic Survey Research Institute, Qatar University*

John Lee Holmes, *Social and Economic Survey Research Institute, Qatar University*

Mohammad Nizam Khan, *Social and Economic Survey Research Institute, Qatar University*

One of the main benefits of using computers in survey research is the integration of randomization in survey designs. For example, in the split ballot technique, the sample is randomly divided equally into two groups and each group receives different questions. Also, in questionnaire design, randomization is needed to control order effects through questions and answer choices rotation.

These techniques are based on a random number generating process whereby the computer generates random numbers which then determine the flow of the interviews. For example, in the split ballot technique, the computer generates a path of random numbers between -1 and 1, and then each respondent is assigned to one of the two groups based on the random number being less than or greater than 0. The shortcoming of this process is that there is no control over the realization of the random numbers. As a result, in any given survey, researchers may end up with significantly uneven groups. In a sample of 1000 respondents, therefore, the randomization may potentially yield 450 respondents in one group and 550 in the other. To address this shortcoming, one common practice is to control the size of the groups by imposing some quotas. However, while controlling for the size of the group, this quota system disrupts the randomization process.

In this paper, we propose a randomization method that is absolutely random and still allows for some control over the realization of the random numbers. The method is based on the antithetic random number generation process developed in Statistics. Instead of generating one path of random numbers for the whole sample, we generate two paths, the primary path and its antithesis. We then apply the method to a CATI survey to demonstrate the effectiveness of the method.

**Saturday, May 17**  
**12:45 p.m. – 1:45 p.m.**  
**Demonstration Session 3**

**A Responsive Web Design for Web Surveys for Smartphones, Tablets and Desktops**

Alerk Amin, *RAND*

Respondents are increasingly using smartphones and tablets to answer web surveys, but then encounter problems when the survey has been designed for a desktop web browser. When respondents cannot read the full question text or answer options, this can lead to inaccurate responses. If the respondent gets frustrated with the experience, this can also lead to drop-offs. Resolving these issues creates a larger burden for survey designers and developers to ensure the surveys work properly on a large variety of devices.

Responsive web design principles can be applied to resolve this problem. The principles of “Mobile First” and “Progressive Enhancement” involve first designing a web page for a mobile device with a small screen, then adding functionality as the screen gets larger. Implementing these principles in our survey system, MMIC, has led to faster development time and better survey experiences for respondents. However, researchers must plan question text and responses for a small screen, sometimes raising issues with the survey design.

This presentation covers how we incorporated these principles into the design of our web surveys, and how the developers worked with researchers to address these issues earlier in the survey design process.

**Mobile Case Management for Real-Time Sample Prioritization Using SMARTField**

Daniel J. Friend, *Mathematica Policy Research*  
Jennifer McNulty, *Mathematica Policy Research*  
Tiffany Waits, *Mathematica Policy Research*

To optimize case prioritization, field interviewers were issued tablets configured with SMARTField, a web-based application developed by Mathematica. This application facilitates real-time case management, web-assisted personal interviews, access to locating information, secure data transmission, and storage on in-house servers. Through real-time monitoring, statisticians identified cases requiring prioritization via an indicator that, when enabled, would display a priority flag on the interviewer’s case list. In this presentation, we will discuss how to successfully manage and prioritize cases in real time during an active field period.

**Saturday, May 17**  
**1:45 p.m. – 3:15 p.m.**  
**Concurrent Session I**

## **Survey Research in a Rapidly Changing Regulatory Environment: Speaking IRB as a Second Language**

### **Ethical Principles and Regulatory Requirements in Survey Research**

Jill E. Darling, *VA Greater Los Angeles Healthcare System*

Federal regulations known as the Common Rule govern human subjects research (HSR). These regulations set out guidelines for review and approval of HSR by an Internal Review Board (IRB). Drafted in response to particularly egregious abuses of research participants, the Common Rule requirements for oversight of HSR are based on the findings of a national commission known as the Belmont Report. The report identifies three basic principles that underlie ethical research practice: respect for persons, beneficence, and justice. Organizations that don't have their own IRBs may work with independent or commercial IRBs, which are governed by the same regulations. Surveys that are not anonymous meet the definition of HSR, but it may not always be clear to researchers when their work must be reviewed by an IRB, or where such a review may be obtained. Surveys that work with identifiable human subjects may not be required to undergo IRB review, if they do not meet the technical definition of research, or if they are not legally required to obtain IRB review. It is incumbent on all researchers to ensure that, even if ethical board review is not required, the procedures they follow in the conduct of their research adhere to the principles set out in the Belmont Report. Jill E. Darling, a long time survey researcher who spent 20 years working with the Los Angeles Times Poll, most recently conducts patient and provider health services research for the Department of Veterans Affairs. Speaking from experience with research ethics inside and outside the regulations that require IRB oversight, Jill will discuss the application of ethical principles to survey research, when IRB approval is required, and provide an overview of resources available to researchers who want to ensure that they are conducting ethical research when IRB review is not required.

### **Interpreting Federal Regulations and Guidance Involving Human Subjects Protections in a Rapidly Changing Technological Environment**

Ronald E. Langley, *University of Kentucky*

Ronald Langley, a long-time survey center director and IRB member will discuss regulatory and ethical considerations regarding using the internet and social media sites for conducting research and recruiting research participants. Areas of particular concern are data security when conducting internet research, when informed consent is required while mining data from social media web sites and whether respondents' confidentiality can reasonably be guaranteed for those who participate in internet research. Advances in technology and the ability to link data from different sources make it increasingly difficult to de-identify data in a manner in which it cannot be easily re-identified by third parties. Additionally, data collection from social media sites also raises questions about privacy expectations of site members and what they would consider an invasion of their privacy. It may be the expectations of privacy are different in these venues and that what constitutes 'minimal risk' (no more than you would experience in everyday



life) are not the same as when conducting other types of research.

Finally, in the absence (as yet) of formal guidance from federal agencies such as the US Department of Health and Human Services' Office for Human Research Protections (OHRP), we will review recommendations such as those in the Secretary's Advisory Committee on Human Research Protections' (SACHRP) report "Considerations and Recommendations Concerning Internet Research & Human Subjects Regulations", guidance available from AAPOR's Code of Ethics and other sources, and discuss the responsibilities of researchers themselves for taking ethical considerations into account when using new technologies to recruit subjects or collect data. We will also provide an update on the status of the Advance Notice of Public Rule Making (ANPRM) and the prospects for a Notice of Proposed Rule Making to be published, as well as some of the hurdles that must be overcome for that to happen.

### **Educating Researchers on IRBs**

Diane Burkom, *Battelle Memorial Institute*

Researchers in many disciplines frequently develop research study protocols that rely on collecting data from human subjects. When federally funded, these studies are reviewed by Institutional Review Boards (IRBs) and cannot begin until (often several) IRBs approve the protocol and associated data collection forms. This IRB review can be very helpful to the researcher when viewed as a valuable assessment of the protocol, instead of a regulatory barrier. This relationship between researchers and the IRB can be improved by educating researchers on how different IRBs work, what IRB members are considering when they review submissions, and how this process can benefit everyone. When researchers at Battelle participated in a mock IRB meeting, the ethical principles and underlying federal regulations created to protect human research participants became more real to them. For example, when researchers understood that some simple changes to their study protocol (such as simplifying the language in a consent form, offering to share study findings with participants, or making some survey questions more generic) could change the risk-benefit ratio enough to obtain IRB approval, they were more willing to adapt their materials and procedures. Studies can produce methodologically rigorous results that satisfy IRB requirements when researchers and board members understand each other's perspectives. Diane Burkom, a program manager at Battelle and IRB member for almost 20 years, will present the benefits of educating researchers about the role of the IRB (both how they operate and what board members are considering) and will present some examples of protocol changes that can improve the experience for the study participants while still allowing the study to collect reliable and valid data.

### **Educating IRBs about Survey Research**

Micheline Blum, *Baruch College, CUNY*

Working successfully and collaboratively with IRBs requires mutual understanding. IRB members may be unfamiliar with, or have misapprehensions about, survey research in general and about public opinion surveys in particular. Such misconceptions need to be addressed by the researcher to avoid demands by the IRB that will jeopardize the research or result in an improper determination or disapproval by the board. It is incumbent upon the survey researcher to educate and inform IRBs about current survey research methods, standards and best practices, levels of risk and generalizability as they relate to the legitimate concerns of the IRB. Micheline Blum, director of a university based survey research center, a survey researcher for more than 30 years, and an IRB member for 5 years, will discuss some of the misunderstandings and misconceptions her own center has faced in IRB reviews, using specific case studies. Examples include total misunderstandings of sampling methodology, data

collection modalities, data security, and best practices as well as the mistaken beliefs that public opinion surveys are not research and that asking sensitive questions poses greater than minimal risk to respondents. Researchers can avoid misunderstandings and the potential negative consequences for their research by understanding IRB concerns stemming from a lack of familiarity and misconceptions as well as from the legitimate concerns of the IRB. The researcher can then clarify the purposes, methodologies, standards and specific procedures of the survey and explain in the IRB protocol how the research addresses the ethical principles, risks and benefits to respondents and participant protections of concern to the IRB.

## **Health Survey Research Methods II**

### **Surveillance of Seasonal Influenza Vaccination Coverage Among Health Care Personnel in the United States**

Charles DiSogra, *Abt SRBI*

Sarah W. Bell, *Abt Associates*

Carla Black, *Centers for Disease Control and Prevention*

Stacie Greby, *Centers for Disease Control and Prevention*

Xin Yue, *Centers for Disease Control and Prevention*

Anup Srivastav, *Centers for Disease Control and Prevention*

Sara M.A. Donahue, *Abt Associates*

David Izrael, *Abt Associates*

Rachel Martonik, *Abt SRBI*

Deborah K. Walker, *Abt Associates*

Influenza vaccination of health care personnel (HCP) may reduce influenza-related illness among HCP and health care settings. Annual influenza vaccination is recommended for all HCP by the Advisory Committee on Immunization Practices. The low prevalence of HCP in the population makes it difficult to sample sufficient numbers of HCPs to obtain information on influenza vaccination coverage during or soon after an influenza season. While national probability-based surveillance systems such as the National Health Interview Survey (NHIS) capture overall HCP vaccination rates, results may not be available for all occupational categories and information on occupational setting and knowledge, attitudes, and behaviors (KAB) specific to HCP are generally not available. While Internet panel surveys can be used to capture HCP information from non-probability surveys, including coverage for multiple occupational categories, occupational settings, and KABs, the reliability of the vaccination coverage estimates from non-probability samples is not known.

During the 2011-12 through the 2012-13 influenza seasons, we used Internet panel surveys that recruited HCP from large web-based panels to monitor influenza vaccination coverage and related KABs. The approach enabled the Centers for Disease Control and Prevention to monitor HCP influenza vaccination coverage and report interim results early during the same season, and final season results before the start of the next season. Our paper describes the final overall influenza vaccination coverage estimates from the Internet panel surveys to those from the NHIS, as well as present influenza vaccination coverage by occupational category, occupational setting, and KABs from the Internet panel surveys. We will describe the strengths and limitations of the different surveillance approaches and the need to validate data from non-probability samples with data from probability samples. This paper will also discuss implications of using non-probability Internet panel survey findings for future monitoring of influenza vaccination coverage among HCP populations.

## **Estimating Population Health in Selected Geographic Areas: Applying Machine Learning Algorithms on Large-scale Survey Data**

LinChiat Chang, *LinChiat Chang Consulting, LLC*

The National Health Interview Survey (NHIS) is a highly respected source of estimates on Americans' health status, health behaviors, and access to health care. Although invaluable, the NHIS data provides estimates mainly at the national level and the 4 census regions - Northeast, Midwest, South, and West. In contrast, many providers of healthcare services do not cover the entire nation or geographies that align well with census regions, thus the current project sought to generate estimates of 3 key health indicators for geographies that could be better aligned with service provider footprints. To do so, machine learning algorithms were applied on high quality large-scale survey data from the NHIS and the U.S. Census Bureau's American Community Survey (ACS). Models were built to predict 3 key health indicators - namely, (a) self-assessed personal health status, (b) frequency of physical exercise, and (c) whether one could not afford medical care when needed during the past year. First, a thorough review was conducted to identify pertinent socio-demographic variables that are available or could be derived from both NHIS and ACS data, including age, gender, race and ethnicity, household income, presence of children, education, occupation, etc. Next, classifiers were built using algorithms such as SVM, neural networks, random forests and Naive Bayes. A majority voting scheme was implemented to combine the results of various classifiers to form a final prediction. The best performing algorithm was then applied on massive ACS data to predict the status of ACS respondents on each of the 3 health indicators. A few case studies will reveal how specific geographic areas of interest can then be delineated with Public Use Microdata Areas (PUMAs) codes in the ACS data. The feasibility and limitations of this approach will be discussed, and comparisons to BRFSS estimates made where applicable.

## **Using “Fit for Purpose” to Design Survey Strategies: Steps Toward Application with Examples from Public Health**

James A. Singleton, *ISD/NCIRD Centers for Disease Control and Prevention*

Mark Messonnier, *ISD/NCIRD Centers for Disease Control and Prevention*

Kirk Wolter, *NORC at the University of Chicago*

Evidence-based decision making has become increasingly challenging in an era of declining public participation in probability surveys. Policy and program decisions may need to be made based on multiple, imperfect data sources. In the 2013 Report of the AAPOR Task Force on Non-Probability Sampling, authors concluded that “Fit for purpose is an important concept for judging survey data quality, but its application to survey design requires further elaboration.” With this broader concept of survey data quality in mind, designers of surveys may be faced with making trade-offs between desired attributes of estimates (e.g., accuracy, cost, timeliness, relevance, and credibility). Evaluation of immunization programs through vaccination coverage assessment provides several examples of this: need for timely influenza vaccination coverage estimates from one season before the next season begins; need for weekly estimates of influenza vaccination coverage in an emergency situation (influenza pandemic or influenza vaccine shortage); and identifying local areas with low vaccination rates for further follow-up based on multiple data sources (small area estimates from surveys and data from Immunization Information Systems or medical claims data). In this paper, we outline potential steps for designing a surveillance system that takes into account desired attributes of the system based on how the estimates will be used. As a first step, we consider the need to delineate how estimates from the system will be used, what actions may be taken based on signals from the

system, and the consequences of these actions (e.g., costs and benefits of accurate, false positive and false negative signals). Subsequent steps involve identification of key attributes of the survey system and establish acceptable trade-offs among these attributes. In some situations it may be possible to quantify impact of these trade-offs using decision analysis. We illustrate this process for several examples from immunization program assessment.

### **Sexuality and Item Nonresponse: Evidence from 12 Years of the California Health Interview Survey (CHIS)**

Matt Jans, *UCLA Center for Health Policy Research*

Ninez Ponce, *UCLA Center for Health Policy Research*

Susan Cochran, *UCLA Fielding School of Public Health*

Annie C. Lee, *UCLA Department of Sociology*

Joe Viana, *UCLA Center for Health Policy Research*

Sexual orientation and sexual behavior are considered some of the most sensitive topics a survey can ask, yet the overall item nonresponse rate is often low (about 2% overall). Low missing rates do not, however, preclude nonresponse bias, particularly if different types of respondents disproportionately do not answer. At minimum, item nonresponse reduces the analyzable data available and introduces possible bias in study findings, which can complicate LGB research on racial, ethnic, and linguistic minorities. The California Health Interview Survey (CHIS), a RDD survey of the California household population, has asked respondents about their sexual orientation since 2001, providing an excellent data source to evaluate changes in sexual orientation item nonresponse over time and explore differences in item nonresponse rates among racial, ethnic, and linguistic minorities. Variation in item nonresponse is explored across five languages (English, Spanish, Chinese, Korean, and Vietnamese), levels of English proficiency, and respondent gender. Nonresponse rates range from below 2% for White respondents and respondents of other races who have high English proficiency, to greater than 5% for Hispanic respondents with low English proficiency or who speak only Spanish. High rates of item nonresponse are also found in some Asian subgroups (e.g., Cambodian and Southeast Asian) but not others (e.g., Korean). Further, there are gender differences in item nonresponse rates within some races/ethnicities, but not within others. Finally, refusal and “don’t know” rates, two components of overall item nonresponse, vary by the gender, race/ethnicity, and language ability of the respondent as well. Overall, preliminary analyses show that there is a complex relationship between respondents’ gender, race/ethnicity, and language, and their propensity to respond to a question about sexual orientation. This suggests that cultural orientation or translation may play a role in item nonresponse for sexually-sensitive questions.

## **Feasibility of Using Immunization Information Systems as a National Immunization Survey Sampling Frame for Monitoring Vaccination Coverage in the United States - Initial Findings**

Stacie M. Greby, *Centers for Disease Control and Prevention*

Sari Schy, *NORC at the University of Chicago*

Andrea Mayfield, *NORC at the University of Chicago*

Margrethe Montgomery, *NORC at the University of Chicago*

Vicki Pineau, *NORC at the University of Chicago*

Laura Pabst, *Centers for Disease Control and Prevention*

LaTrece Harris, *Centers for Disease Control and Prevention*

Larry Wilkinson, *Centers for Disease Control and Prevention*

James A. Singleton, *Centers for Disease Control and Prevention*

The National Immunization System (NIS) is a dual-frame landline/cell telephone survey used by the Centers for Disease Control and Prevention (CDC) to monitor vaccination coverage among children 0-17 years. A substantial portion of data collection costs are for initial screening for households with children aged 19-35 months. Low response rates have increased survey costs. Immunization Information Systems (IIS), also known as registries, exist in 49 states, five large cities, and the District of Columbia. IIS enroll children at birth and update vaccination information through provider-reporting. Substantial efficiencies could be gained by using IIS to supplement or replace the current NIS sampling frame, provided an IIS is mature enough to capture reliable contact information for use by the NIS. The use of IIS as a sample frame for the NIS is being assessed in IIS with varied capacity. This paper summarizes initial findings about the feasibility of using IIS as an NIS sample frame from both the NIS and IIS perspectives. Analyses include comparisons of key telephone data collection rates in the IIS and NIS samples, including contact phone number rates, eligibility rates, household screener and interview completion rates, and consent rates to contact providers and the IIS for children 19-35 months in the household. We also discuss challenges and solutions for meeting varying state and federal human subject protection requirements and differing procedures surrounding the exchange of vaccination-related information and contact information needed for fielding data collection. Additionally, we discuss cost modeling, timing, and IIS resources that could impact adoption of IIS sample as a stand-alone/supplemental sample source in the NIS. Finally, we discuss the potential impact of utilizing data from IIS on the comparability of NIS vaccination coverage across state and local levels due to the different IIS data collection sources and data processing methodologies.

## **Women's Attitudes in Conflict, Crisis & Post-War Zones: Research from Public Opinion Surveys**

### **Afghan Perspectives on Taliban Reconciliation**

John Richardson, *D3 Systems, Inc.*

The question of Taliban participation in the Afghan government after the majority of NATO troops leave Afghanistan in 2014 is one of the most vital to the country's future. While many Afghans welcomed the relaxation of restrictions on women's rights after the fall of the Taliban government, the gains Afghan women have made since 2001 will be contested when Taliban members and other conservative elements that had opposed the NATO presence in the country are integrated into the government. Since 2010, ACSOR-Surveys has been collecting public opinion data in Afghanistan regarding reconciliation between the Taliban and the central

government and changes to women's rights that Afghans would view as acceptable conditions for a stable government that included Taliban members. This paper, using longitudinal data from the Afghan Futures survey conducted by D3 Systems, Inc. tracks attitudes towards reconciliation between the Taliban and the central government, with special focus on the views of Afghan women.

### **Perceptions and Predictors of Inequality Among Afghan and Pakistani Women**

Melissa Abadi, *Pacific Institute for Research and Evaluation*

Stacey Frank, *D3 Systems, Inc.*

Stephen Shamblen, *Pacific Institute for Research and Evaluation*

Samantha Lee-Ming Chiu, *D3 Systems, Inc.*

Kirsten Thompson, *Pacific Institute for Research and Evaluation*

The people of Afghanistan and Pakistan have endured over 30 years of oppression, poverty, violence, and widespread human rights violations. While living in war-conflicted countries takes its toll on everyone, the women in these countries also bear extreme gender-based restrictions which often deny them access to healthcare services, education, employment, and political participation as well as basic human rights. Despite these restrictions, women are integral to achieving and sustaining peace in these countries and are making initial strides to protect their rights and freedoms. Our paper will assess perceptions of inequality among Afghan and Pakistani women as well as the strongest predictors of perceived inequality, including access to healthcare services, education, employment, political participation, legal access, and protection from violence through laws. We will also explore the impact of different media sources on perceptions of inequality in order to examine whether access to and use of Western media is associated with perceived inequality among these women.

This paper will utilize data collected in Pakistan and Afghanistan in 2012. Both are face-to-face surveys of nationally representative samples drawn using a stratified, multi-stage selection procedure. The Pakistani survey was conducted among a random sample of 2,012 adults from March 19th-30th, 2012, though this paper will focus on the opinions of the 968 female respondents. The sample included the four Pakistani provinces and Islamabad Capital Territory, and excluded the Federally Administered Tribal Areas. The Afghan survey was conducted from January 24-February 3, 2012 among a random sample of 2,018 Afghan adults, with 971 female respondents. Women were interviewed in all Afghan provinces other than Paktika, which was inaccessible at that time for female interviewers.

### **Gender Differences: Public Opinion During the 2013 Political Crisis in Egypt**

Nina Sabarre, *D3 Systems, Inc.*

Elizabeth Hood, *D3 Systems, Inc.*

David Rae, *D3 Systems, Inc.*

Since the ousting of Egyptian President Mohamed Morsi on July 3rd, 2013, the country remains in a state of uncertainty. Supporters of the military's intervention claim the coup was necessary to prevent civil war and restore democracy. Months later, however, it is clear that the military is failing on both accounts. Hundreds of protestors have been killed and the political future of Egypt remains volatile. Although the divide between Islamists and secularists is most frequently discussed, differences among key demographic groups also explain Egyptians' attitudes and behaviors towards the political turmoil. Using survey data collected in June 2013, this study investigates public opinion of the current political situation in Egypt, specifically focusing on key differences between men and women.

D3 Systems, Inc. commissioned a national CATI-based survey consisting of 1,001 Egyptian adults, split into two phases of approximately 500 interviews each. The purpose of the two-phase study was to experiment with probability-based sampling approaches. Phase I was completed between June 27th and July 1st. On July 3rd, the Egyptian military removed President Morsi and announced the installation of a new civilian government. Fieldwork for Phase II was completed between July 5th and July 8th. Although this two-phase study was designed as a methodological experiment, the timing of the military coup allows for opportune analysis of public opinion before and after the removal of President Morsi.

The authors of this paper analyze shifts in opinions from Phase I to Phase II, to understand how the military coup may have influenced the views of men and women in Egypt. This paper supports the hypothesis that Egyptian women were less optimistic in the days following the coup than Egyptian men, as women remain skeptical of progress made after the 2011 revolution.

### **Women, Peace and Security in a Nascent State: A Case Study on South Sudan**

Mayesha Alam, *Georgetown University, Institute for Women, Peace and Security*

Rebecca Turkington, *Georgetown University, Institute for Women, Peace and Security*

Brian Kirchhoff, *D3 Systems, Inc.*

Through the Women In Muslim Countries (WIMC) research, we have collected data on Muslim women across the globe. From 2007 to 2013, WIMC collected data on 25 countries: Afghanistan, Bahrain, Bangladesh, Bosnia-Herzegovina, Egypt, France, Lebanon, Oman, Pakistan, Qatar, Saudi Arabia, Spain, Iraq, Iran, Italy, Jordan, Kazakhstan, Kosovo, Kuwait, Kyrgyzstan, Syria, Tajikistan, Turkey, UAE, and Yemen. Countries are continuously added. Most countries have national stratified random samples to capture the attitudes and behaviors of the female population. We measure women's empowerment in actual daily practice, providing a deep look into the oft-perceived gap between current public policy and empowerment initiatives and actual practice on the personal and local level. Key findings and cross-national trends and comparisons across countries are examined. We report significant differences in opinion and/or behavior between demographic groups (age, education, employment, ethnicity and income). The survey results suggest a rich and diverse spectrum of opinions amongst Muslim women. The complex methodology and comparisons between face-to-face and CATI research designs are discussed.

### **Women In Muslim Countries: A Multinational Analysis**

Samantha Lee-Ming Chiu, *D3 Systems, Inc.*

Dameka Williams, *D3 Systems, Inc.*

Survey research has a role to play in providing women with a public voice where custom and culture may not permit them their own. As part of a collective effort to understand the status of rights (in practice and beliefs) of Muslim women around the world we have collected data from 25 countries in the Women in Muslim Countries (WIMC) Project. The program is designed to measure women's empowerment in actual daily practice, providing a deep look into the oft-perceived gap between current public policy and empowerment initiatives and actual practice on the personal and local level.

Through the WIMC program, D3 Systems has collected data on Muslim women across the globe. From 2007 to 2013, WIMC collected data on 25 countries: Afghanistan, Bahrain, Bangladesh, Bosnia-Herzegovina, Egypt, France, Lebanon, Oman, Pakistan, Qatar, Saudi

Arabia, Spain, Iraq, Iran, Italy, Jordan, Kazakhstan, Kosovo, Kuwait, Kyrgyzstan, Syria, Tajikistan, Turkey, UAE, and Yemen. Countries are continuously added. Most countries have national stratified random samples to capture the attitudes and behaviors of the female population as a whole.

This paper highlights the complex methodology and results of this multinational survey. Key findings and cross-national trends and comparisons across countries are examined. This paper reports significant differences in opinion and/or behavior between demographic groups (age, education, employment, ethnicity and income). The survey results suggest a rich and diverse spectrum of opinions amongst Muslim women. The complex methodology and comparisons between face-to-face and CATI research designs are discussed.

## **Qualitative Research: The Design & Implementation of Qualitative Research Approaches that Add Value to Mixed-Methods Designs**

### **Designing and Implementing Qualitative Approaches in Mixed-Methods Evaluations of Enumeration and Coverage in Decennial Censuses**

Laurie Schwede, *U.S. Census Bureau*

Rodney Terry, *U.S. Census Bureau*

An ongoing challenge of the Census Bureau is persistent differential undercounts of some minority groups across decennial censuses, including the 2010 Census, despite continued efforts to reduce them. For the 1990 and 2010 Censuses, Census Bureau ethnographers designed and conducted mixed methods qualitative/quantitative evaluations. In both evaluations, ethnographers conducted research in households in race/ethnic sites to decide who should and should not be counted there, who may have been missed, and why errors may have occurred. Those households and names were then matched to final census records in partial quantitative validation studies to identify the factors that may have led to miscounts and who may have been missed. The qualitative methods differed: in 1990, ethnographers conducted independent local censuses separate from the census, while in 2010, ethnographers accompanied census interviewers to observe live interviews and debrief respondents. This pair of evaluations with the same objectives but with different qualitative methods offers us a unique opportunity to address the overall theme and outline of the proposed AAPOR session, “Qualitative Research: The Design and Implementation of Qualitative Research Approaches that Add Value to Mixed-Methods Designs.” In this paper, we compare and contrast the qualitative design and implementation features of the 1990 and 2010 mixed methods evaluations in terms of 1) methods and modes; 2) design features; 3) analysis (process and method); 4) contribution of the qualitative research to advancing understanding of who is miscounted and why and 5) the core importance of the ethnographers’ contributions in generating the lists of households and persons for the partial validation study with the final census records. We analyze the strengths, limitations, and challenges of the two qualitative approaches and conclude with recommendations for improved mixed qualitative/quantitative methods for a future 2020 Census evaluation on who is miscounted and why.

### **The Study of Construct Validity in Survey Estimates: Expanding the Role of Cognitive Interviewing to Reveal the Meaning Behind the Statistic**

Stephanie Willson, *National Center for Health Statistics*



Traditionally, cognitive interviewing has been used to pretest survey questions for problems that might elicit response error, which in turn can produce faulty estimates. Cognitive interviewing was seen predominately as a question design tool. This paper argues that the method can offer more.

Cognitive interviewing can be used in a mixed-method endeavor with survey research to study the construct validity of a survey question and to shed light on the meaning behind a survey statistic. This paper will illustrate, through example, how question meaning is studied through qualitative methodology. First, it argues that it is only through qualitative interviewing that we hear respondents describe, in their own terms, their interpretation of a survey question. These interpretations are the building blocks of the construct being measured. Additionally, constructs are often convoluted concepts, impossible to capture in a single question. When survey questions cannot function identically for all respondents, the results of a cognitive interview study can benefit data users who want to gain insight into the true meaning of the statistics they use. It is in this capacity that cognitive interview study results work in a complementary fashion with quantitative data. This is an underutilized, but invaluable, aspect of the cognitive interviewing method.

An example is drawn from a study being conducted by the National Center for Health Statistics on the Donor Risk Assessment Interview. This is a questionnaire used to evaluate organ and tissue suitability from potential donors. The questions are administered by proxy to a loved one of the deceased potential donor and focus on the health conditions and behaviors of the deceased. For example, this study has found that the nature of the death experience – that is whether it was short and sudden or long and drawn out – has an effect on question interpretation.

## **Using Qualitative Data to Develop, Supplement and Illustrate Quantitative Survey Research**

Kinsey Gimbel, *Fors Marsh Group*

Jennifer Romano Bergstrom, *Fors Marsh Group*

Researchers sometimes see qualitative methods as something to be used only in the very initial stages of a project, as techniques to be left behind once a quantitative survey method begins. However, qualitative techniques, including focus groups and in-depth interviews, can both supplement survey findings and provide unique data unobtainable through quantitative methods. This paper will use an on-going study to show how integrating qualitative methods throughout the research process results in more complete, powerful data.

Specifically, we will describe how focus groups have been used throughout the project, from survey development to data presentation, including:

- During the initial survey creation and during revisions of the survey for subsequent fieldings, focus group data was used to evaluate prospective survey questions, identify possible response options, and refine question wording.
- When identified areas of interest were difficult, if not impossible, to address in a survey format, focus groups were conducted to collect these data.
- During presentation of the data, quotes and video clips from focus groups were used to illustrate quantitative findings and ensure that the final results and recommendations to the client have maximum impact.

The goal of the paper is to provide guidance on how qualitative data can be integrated into a

larger quantitative project to provided added value and offer alternate ways to address research questions. Additional areas of discussion will include how the focus groups were structured to ensure consistent, quality data; how qualitative data was coded and analyzed; how qualitative data was integrated with quantitative survey findings; and how findings were presented in usable, actionable format.

## **Benefits and Challenges in Conducting Web-Mediated Qualitative Interviewing**

Vince Welch, Jr., *NORC at the University of Chicago*

The benefits to coverage, reliability, and validity of qualitative data collection (e.g., expert panels, focus groups, and cognitive interviews) in the design phase of questionnaires is well documented (Campanelli, 2008; Willis, 2004). Qualitative interviews are generally conducted in a face-to-face setting which can lead to extensive costs associated with travel and difficulty in scheduling. Due to the time and monetary costs associated, the overall number of interviews in a qualitative research program tends to be quite low.

In an effort to overcome some of the cost-related drawbacks of qualitative data collection, researchers have explored electronically mediated methods of qualitative data collection. Noel (2013) explored conducting cognitive interviews over the telephone. Edgar (2011, 2013) explored the utility of web-based, un-moderated cognitive interviews for questionnaire testing. While these methods showed some promise, the inability to record and respond to non-verbal cues over the telephone or to insert additional probes in Edgar's (2013) web-based data collection were significant drawbacks.

The near-ubiquity of high-speed web access and web cameras in some segments of the population presents researchers with an opportunity to have widely geographically dispersed respondents in the same virtual 'space' to conduct qualitative interviews. The current work explores the utility of using web-based group video conferencing (e.g., Google +; Skype) to conduct qualitative interviews. We are conducting a series of cognitive interviews and focus groups using Google + 'hangouts' in a population of recent doctorate recipients and graduate school administrators. We chose the Google + platform because it offered the ability to simultaneously share data (e.g., survey forms), the ability to automatically record and share, and zero platform-related cost. We present insights into the logistics, benefits, and challenges of conducting focus groups and cognitive interviews using web-based, multi-way audio/video/data platforms.

## **Probability Sampling for Qualitative Researchers: A Flexible Strategy**

Su Li, *University of California-Berkeley Law School*

William Welsh, *University of California-Berkeley Dept. of Sociology*

Qualitative research, especially that which uses in-depth interviewing to collect data, has made immense contributions to social science and in applied areas. However, the nonprobability sampling design of these studies leaves them open to criticism over generalizability and reproducibility (e.g. Lucas 2012). Such research can be made even more powerful if practitioners adopt probability sampling in research design and data collection. This paper presents a novel, flexible, yet rigorous way to draw stratified probability samples for qualitative research, both academic and applied. The procedure generates a ranked list of units and correctly assigns each a sampling weight. Crucial for the often contingent nature of qualitative research, our approach allows the researcher to modify the sample mid-project without compromising its integrity.

We first discuss examples of qualitative work whose generalizability and reproducibility could have been strengthened with the use of our approach. We review data collection designs of qualitative work published in major sociology and social-legal research journals to highlight the lack of rigorous sampling design. After a brief formal justification, we provide step-by-step instructions. We then apply our technique to both simulated and real-world data. The real-world data comes from a mixed method empirical project on law firms. We explain in detail how we sampled 80 law firms out of 176 law firms using the stratified probability sampling method. Strata are defined using variables that we locate at the research design stage. Descriptive statistics and significance tests are presented to illustrate the difference of research results with or without correctly incorporating the sampling weight. Most important, our sampling strategy provides high-level flexibility for qualitative researchers to adjust their sample size dynamically onsite when they are in the field and encounter the issue of “no access” or “nonresponse”. Lucas, Samuel Roundfield. 2012. “Beyond the Existence Proof.” *Quality & Quantity*, 1–22.

## **Interviewers and Interviewing: Rapport, Respondent Engagement and Sensitive Behaviors**

### **Face-to-Face vs. Video-Mediated Interviews: Rapport Between Interviewers and Respondents and the Disclosure of Sensitive Information**

Hanyu Sun, *Joint Program in Survey Methodology*

Frederick G. Conrad, *University of Michigan*

Frauke Kreuter, *University of Maryland*

Technological advances in recent years have made video-mediated interviews more feasible and affordable; however, little attention has been paid to videoconferencing as a potential mode of data collection. In video-mediated interviews, the interviewer and respondent can see and talk to each other via a video window. Video-mediated interviews provide several potential advantages for surveys. Respondents of video-mediated interviews may feel more engaged or connected than those in telephone interactions due to a greater sense of social presence. This may lead to higher completion rates and better data quality. It is a cost-saving alternative to in-person interviews, especially when interviewing geographically dispersed respondents. Additionally, there may be certain types of questions that especially benefit from social distance through video-mediated interviews instead of face-to-face interactions. However, these hypotheses have so far not been tested empirically. Although rapport has been found to increase disclosure of moderately sensitive information in face-to-face interaction (van der Zouwen, Dijkstra, and Smit, 1991), it is unknown if rapport can be established to the same extent in video-mediated interviews leading to similar levels of disclosure. We compare video-mediated interviews with face-to-face interviews in a laboratory experiment to investigate (1) whether rapport can be similarly established in video-mediated and computer-assisted personal interviews (CAPI); and (2) whether video-mediated interviews increase disclosure of moderately sensitive information to the same extent as CAPI. The experiment varies both the level of rapport (high-rapport vs. low-rapport interviewers) and the mode of data collection (CAPI vs. video-mediated interviews). Eight professional interviewers and 128 respondents will participate. Adobe Connect will mediate the video interviews, allowing showcards to be displayed remotely. We will compare sensitive admissions as well as rapport ratings between CAPI and video-mediated interviews. We will present findings and discuss their implications for future survey data collection.

## **Moral Exemplars, Outpatient Medical Clinic Climate, Temporal Affect and Patient Care Errors**

Deborah M. Mullen, *Park Nicollet Institute*

The Outpatient Medical Clinic Safety Climate was created to evaluate the relationship between the rate of patient care errors, the clinic climate in outpatient medical practices and health care provider personality and temporal affects. The surveyed population included every Nurse Practitioner and Certified Nurse Midwife holding 2009 Minnesota licensure. The instrument was administered through a mail, return mail process to 2,576 professionals resulting in a 52% response rate (AAPOR RR1). Survey dimensions included: reported error rates, clinic climate and culture, as well as individual respondent's temporal affect and personality traits (moral exemplarism).

Latent variable development focused on identification of climate, culture, moral exemplarism, and temporal affect. Exploratory factor analysis allowed for the grouping of survey items into scale scores. Generalized linear modeling was utilized for final modeling. Final models included separate models for personal and clinic errors reported.

Findings: For personal errors, those made by the respondent, the presence of a safety climate and a medication reconciliation process increased the number of reported errors. For those errors made in the clinic (clinic errors, medical errors, adverse events, near misses, and accidents waiting to happen) multiple culture scales were significant as well as the existence of a safety climate. Culture scales: formal communication about safety, error reporting process and just culture as well as safety climate all correlated with increased reported clinic error. Temporal affect – causal beliefs and moral exemplarism scales were not found to be meaningful contributors to any of the models. Clinic errors, medical errors, adverse events, near misses, and accidents waiting to happen are strongly related to by the clinic's culture and climate. As climate and culture are shared perceptions, then it seems reasonable that for the clinic as a whole these factors would explain more of the error model.

## **The Role of Rapport and Standardization in Predicting Future Survey Participation: Interviewer-Respondent Interaction in Questions about End-Of-Life Planning and Preferences**

Dana Garbarski, *University of Wisconsin-Madison*  
Nora Cate Schaeffer, *University of Wisconsin-Madison*  
Jennifer Dykema, *University of Wisconsin-Madison*  
Ellen Dinsmore, *University of Wisconsin-Madison*  
Bo Hee Min, *University of Wisconsin-Madison*

"Rapport" is a vague concept that has been used to refer to a wide range of features of interaction. Rapport is sometimes assumed to be a positive psychological feature of the interaction, referring to a situated sense of affiliation between interactional partners, comfort, willingness to disclose or share, motivation to please, or empathy. Rapport may benefit survey response quality by increasing respondents' motivation to disclose or provide accurate information, but could also harm data quality if motivation to please or affiliate causes respondents to suppress undesirable information.

Questions about one's end-of-life planning and preferences are potentially sensitive and interactionally delicate for both interviewers and respondents, creating a unique opportunity to study the development and maintenance of interactional rapport. Unlike earlier studies, we

examine rapport as an interactional phenomenon, describing particular qualities of the interviewer-respondent interaction rather than focusing on the psychological dimensions of rapport. We define dimensions of rapport in interaction both in terms of the content of what is said as well as the sequencing or structure of talk. Given that behaviors in the survey interview are constrained by the rules of standardization, we attend to the opportunities for creating and maintaining rapport within these constraints.

Using transcripts from audio recordings of the end-of-life section of the 2004 wave of the Wisconsin Longitudinal Study, we observe that within the interview, rapport appears in responsive behaviors by interviewers and behaviors suggesting engagement by respondents. We discuss contexts in which behaviors oriented toward rapport conflict or complement behaviors mandated by standardization.

Rapport developed during the interview may affect respondents' decisions to participate in subsequent data collection efforts. Using a case-control design, we examine whether interviewers' responsiveness and respondents' engagement measured during the 2004 end-of-life module predict respondents' likelihood to participate in a subsequent request to provide a saliva sample.

### **Interviewer Variance and Prevalence of Verbal Behaviors in Calendar and Conventional Interviewing**

Robert F. Belli, *University of Nebraska-Lincoln, Survey Research and Methodology Program*  
Nuttirudee Charoenruk, *University of Nebraska-Lincoln*

Past work in retrospective reporting in surveys has looked at calendar and conventional questionnaire interviewing in regards to interviewer variance on responses (Sayles, Belli, & Serrano, 2010) and verbal behaviors among respondents and interviewers (Bilgen & Belli, 2010), but interviewer variance on verbal behaviors has not yet been examined. In this research, we compare the interviewer variance of respondent and interviewer verbal behaviors between 165 calendar and 162 conventional interviews. The types of interviewer behaviors that were examined include those for retrieval probes, standardized associated behaviors, conversational behaviors, feedback, and rapport. The types of respondent behaviors examined include the use of retrieval strategies, probe disagreements, expressions of cognitive difficulty, conversational behaviors, and rapport. We find that almost all of interviewer variances of the behaviors differ significantly between calendar and conventional methods, the lone exception is for respondent laughter. The interviewer effects in all remaining respondent behaviors in calendar interviewing is higher than in conventional interviewing; however, the size of the interviewer effects for interviewer behaviors between methods are not in the same direction. By relating the interviewer variances on key verbal behaviors to their prevalence between the two interviewing methods, we infer whether the interviewer effects on the behavior were driven by the methods requiring different styles of interviewing, or by the interviewers themselves.

## **Cultural Variations in the Effect of Interview Privacy and the Need for Social Conformity on Reporting of Sensitive Outcomes**

Zeina N. Mneimneh, *Institute for Social Research, University of Michigan*

Mick Couper, *Institute for Social Research, University of Michigan*

Roger Tourangeau, *Westat*

Beth-Ellen Pennell, *Institute for Social Research, University of Michigan*

Steven Heeringa, *Institute for Social Research, University of Michigan*

Michael Elliott, *Institute for Social Research, University of Michigan*

The interview setting is one of the important factors that could affect the measurement properties of the data collected. Yet, it is a factor that has not been well-studied and understood. Surveys that are designed to collect sensitive information usually call for a private interview setting. In reality, interviewers are essentially guests in respondents' homes and find it difficult to achieve privacy. Thus, a substantial proportion of interviews are conducted in the presence of a third party even when the study protocol calls for privacy.

The literature on the effect of third party presence on reporting sensitive information is mixed. While some studies found that third-party presence is associated with reduced reporting of sensitive outcomes others observed increased reporting of such outcomes. This inconsistency could be attributed to a number of factors including: inconsistent definitions of third-party presence and duration of presence across studies, differences in controlling for possible confounding factors, failure to investigate respondent characteristics that could moderate the effect of third party presence on reporting, and cultural differences across studied populations. To address these factors, survey data from a cross-national initiative that uses comparable survey design and implementation procedures is analyzed. While controlling for possible confounding variables and interviewer variation across these surveys, we investigate the moderating effect of respondent's need for social conformity and respondent's cultural background on third party presence and reporting of sensitive outcomes. We also contrast reporting patterns for sensitive outcomes with those for a number of non-sensitive outcomes. Findings reveal that while third party presence affects the reporting of sensitive outcomes, it does not affect the reporting of non-sensitive outcomes. Third party presence effects are moderated by the respondent's need for social conformity and the respondent's cultural background. However, these moderating effects differ depending on whether the outcome is behavioral or attitudinal.

## **Explorations of Muslim Public Opinion**

### **After the Coup: Egyptian Public Opinion in the Summer of 2013**

Samuel Solomon, *D3 Systems, Inc.*

Alex Brezinski, *D3 Systems, Inc.*

In the final weeks of June 2013, millions of protesters amassed in Cairo and all across Egypt to protest the government of Muhammed Mursi. Within days, the armed forces arrested the president, suspended the constitution, and presented a "road map" to a new political future for Egypt. While a coterie of military, civilian, and religious figures appeared on television to bless the July 3rd coup, public support was evident from broadcasted images of the celebrating crowds around Egypt.

The events of summer 2013 underlined the rising significance of public opinion in post-Mubarak Egypt. Even as Egypt's political transition has taken sharp and sometimes unexpected turns,

public opinion has continually informed the decisions of Egypt's political and military leadership, before and after July 3rd. This paper examines the views of Egypt's population in the days leading up to and the days following the July 3rd coup. The authors investigate the political attitudes of Egyptians across a wide cross-section of demographic variables – including geography, age, education level, income, and occupational status – to analyze who supported and who opposed the actions taken by the armed forces in July 2013.

This paper utilizes data from a nationally representative CATI survey of 1,001 Egyptians commissioned by D3 Systems, Inc.. The first half of fieldwork for the survey was conducted from June 27th to July 1st, while the second half was completed from July 5th to July 8th. The timing of the fieldwork thus allows for pre- and post-coup comparisons of Egyptians' political opinions.

### **Islam in Central Asia: An Application of LCA to Study Patterns of Religious Observance**

Igor Himelfarb, *Educational Testing Service*

Neil Esipova, *Gallup Inc.*

In 2008, Gallup as a part of World Poll data collection efforts, included items probing the respondents' level of religiosity and observance of Muslim religious customs in the country surveys. The participants were asked to rate the degree of practice of main Islamic religious customs. The items on the survey probed the frequency of daily prayer (salat), the observance of Ramadan, the compliance with Islamic dietary laws (halal), pork and alcohol consumption, the desire to perform Muslim pilgrimage to Mecca (hajj), compliance with (for females) and support (for males) for Islamic head cover for women (hijab), acceptance of interfaith marriage, support for establishment of places of worship in public areas, and support for Islamic laws (Shariat) in respondents' country.

To classify the Muslim population living in the Kazakhstan, Kyrgyzstan, Azerbaijan, Uzbekistan, and Tajikistan, a statistical method of latent class analysis (LCA) was used. Five independent LCA models, one for each of the countries, were estimated and tested. This method is a multivariate exploratory analysis that helps to identify latent classes in the population based on binary ("yes"/"no") responses to a survey. LCA clusters the individuals into classes and computes the probabilities of endorsing a survey item (answering "yes" versus "no") conditional on the class membership. First, a one-class, two-class, three-class, four-class, five-class, and six-class models were considered. Then, the patterns of responses as well as numerical fit indices were examined to select the best fitting model.

Similar patterns of religious observance were discovered in all five countries examined. The three-class models revealed the best fit to the data in all countries divulging Muslims who partially practice Islamic faith (Part-Time Pious Class), fully practiced Islamic faith (Devout Pious), and a unique class in which individuals would practice or support some religious customs while eating pork and consuming alcohol (Soviet Muslims Class).

## **News and Entertainment Media Use in the Arab World**

David Krane, *Harris Interactive*

Everette E. Dennis, *Northwestern University in Qatar*

Justin D. Martin, *Northwestern University in Qatar*

Robb Wood, *Northwestern University in Qatar*

Since 2011 interest in the media in the Middle East has intensified with much speculation but relatively little systematic research. At this time of such intense interest Northwestern University in Qatar commissioned Harris Interactive to conduct a large survey project assessing news, entertainment and media use in eight Arab countries (Tunisia, Egypt, Lebanon, Qatar, Saudi Arabia, Bahrain, UAE, and Jordan). Surveys were conducted among nationally representative samples in each country using either face-to-face or telephone interviewing. In total 10,027 interviews were completed (approximately 1,250 per country).

The survey focused on media use with a particular focus on the internet, social media, news, and entertainment information. The questionnaire also included items assessing respondents' perceived levels of political cynicism, political efficacy, online privacy and credibility of their nations' news institutions. In a region where interpersonal and face-to-face communication is especially important we examined how internet use, including social media is employed.

Briefly, amid the rapid adoption of new media across the regions and the relative stability of traditional media, including newspapers and magazines, there is still some ambivalence about the impact and influence of the internet. Most respondents are optimistic about the internet as a medium of personal and political communication with large majorities agreeing that people should be able to express their opinions online, no matter what those opinions might be. That commitment to freedom of expression in the abstract, however, breaks down on closer inspection since majorities in four countries – Lebanon, Saudi Arabia, Qatar and Tunisia – also want tighter regulation of the internet in their country. Those in Bahrain, Jordan and UAE have mixed feelings about increased government control. The paper will explore these and other findings from this project.

## **Does the Quest for Democracy Necessarily Mean the Quest for Gender Equity in the Middle East?**

M. Nizam Khan, *SESRI, Qatar University*

The past few years have witnessed the mass movements against autocratic rule and struggle for democracy in the Middle East-North Africa regions. Active participation of women was quiet noticeable in these movements and apart from the general demands for a more inclusive political society and enhanced socioeconomic justice, many female protesters in these countries, where gender inequality was a prevailing feature, hoped to enhance their status in society. However, the post movement governments appears to be committed more to religious norms than to citizen or women's rights and the post movement period has been marked by marginalization of women and minorities and divisiveness. This raises a very fundamental question, whether gender equality is considered as an essential component of democracy in this part of the world.

The main purpose of this study is to explore the relationship between the quest for democracy and gender equality in the Middle East using survey data.



Data for this study came from the World Values Surveys (WVS) conducted in the Middle East (during 2005-2010) and the 2012 Qatar Omnibus Survey. The countries included in this study are: Egypt, Jordan, Iran, Turkey, and Qatar. In all countries, except Iran (34%), more than half of the respondents believed that it was absolutely important that the country they lived in was governed democratically. When asked how essential the gender equity was as a characteristics of democracy, more than half in Turkey, Egypt and Jordan, and about one-third in Iran and Qatar reported that it was absolutely essential. The attitudes towards democracy and gender equity will be examined by age, gender, education, employment, and economic status.

### **The Effect of Acculturation on Extreme Response Style: A Mediation Analysis Among a Sample of Arab American Adults**

Mingnan Liu, *University of Michigan*

Sunghye Lee, *University of Michigan*

Frederick Conrad, *University of Michigan*

Response style (RS) refers to the fact that when answering Likert-type questions, respondents use certain categories more frequently than others regardless of the question contents. Although numerous studies have examined the RS differences across cultural groups, so far we have limited knowledge about the RS patterns of immigrants, a bi-cultural group who possesses both host and home country cultures. When they participate in surveys, bi-culturals can exhibit different RS.

In this paper, we explore the impact of acculturation on extreme response style (ERS) among Arab American respondents in the Detroit Arab American Study. ERS refers to response behavior in which the two endpoints of a response scale are chosen more often than other categories. Acculturation level is operationalized by years living in the U.S., number of parents born in the U.S., language used at home, watching TV in Arabic regularly, and the frequency of interacting with people of other race/ethnic backgrounds. Using multi-level generalized structural equation modeling, a mediation analysis is conducted where the language of the interview (i.e., Arabic vs. English) serves as the mediator. The preliminary results show that as Arab Americans become more acculturated, less ERS is evident. More specifically, significant indirect effects of all five acculturation measures on ERS were observed. Also, after adjusting for the mediator, partial support of the direct effects of acculturation on ERS was also evident in the analysis. The mechanism behind this is that Arabic culture is predominantly honor-based and collectivist. When these cultural attributes are reflected in RS, we expect to see more ERS since extreme responses are viewed as relatively assertive and are preferred as a way of showing one's unambiguous attitude and standing, an important quality in honor-based society. The survey language mediated the acculturation impacts by priming the relevant culture, which in turn affects the RS.

## **Investigating Public Opinion Trends in the U.S.**

### **Measuring Fear of Leviathan: Cross-Time Changes in the Public's View of the Power of the Federal Government**

Robert W. Oldendick, *University of South Carolina*

Christopher Werner, *University of South Carolina*

Dennis N. Lambries, *University of South Carolina*

The issue of the power of the federal government has been at the core of divisions in American politics since this country's founding, from the debate over the ratification of the Constitution through the Civil War to the rise of the Tea Party. Although recent polls (e.g., Gallup September 2013) have shown that Americans' belief that government has gotten too powerful has reached records levels, there are a number of opinions on specific issues that question this characterization of the public's view. This paper employs data from a number of sources including Gallup, the American National Election Studies, the General Survey Social, and the Pew Research Center in a cross-time analysis that examines the question of how the public's views of the federal government's power are dependent on the way in which this concept is measured. The results show that not only is the federal government power better measured as a multi-dimensional concept, but the nature of these dimensions has changed over time. Significant differences in views of the power of the government are evident among groups, with distinctions by party identification, political ideology, and race of particular importance, although the nature of these differences has also changed over the 50 year period in which opinions on this issue have been measured.

### **Tracking America Across Four Decades: The General Social Survey, 1972-2012**

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

The 29 rounds of the General Social Survey have interviewed 57,061 respondents and have data on 2,072 trends. This paper examines several hundred of those trends and presents an overall profile of societal change during this period. The main topics covered are 1) abortion rights, 2) civil liberties, 3) confidence in institutional, 4) crime and violence, 5) economic well-being, 6) gender roles and gender equality, 7) governmental spending priorities, 8) intergroup relations, 9) psychological well-being, 10) religion, and 11) sexual morality.

### **Empirical Support for the Media Participation Hypothesis: Longitudinal Trends Across Presidential Elections**

Erik P. Bucy, *Texas Tech University*

Jacob Groshek, *Boston University*

This study investigates the media participation hypothesis proposed by Bucy (2005) by examining the use of different media platforms and formats across the six most recent U.S. presidential elections (1992-2012). The hypothesis holds that, as political involvement becomes increasingly reliant on new communication formats and technologies, intensive use of interactive public affairs media will produce a heightened sense of system satisfaction and political efficacy, a trend that should manifest itself longitudinally as mass media become more interactive in nature. Multivariate analyses using American National Election Studies data demonstrate that as participatory media become more prevalent and utilized in an electoral system, corresponding and statistically significant increases in political efficacy and perceptions of government responsiveness can be observed among the heaviest users of participatory media. Enhanced engagement with interactive media, in other words, equates with positive

assessments of democratic processes. At the same time, increased engagement with participatory media also shows an inverse, negative relationship with trust in government, suggesting that the close-up and sometimes overheated view that interactive media offer of politics can have corrosive effects on political attitudes.

Results of the study are interpreted within the context of technology's influence on civic life. With the rise of interactive technology, media use is no longer a simple matter of information seeking and mere exposure but an experience that increasingly resembles civic involvement itself. Indeed, save perhaps for holding office, almost all forms of civic and political participation found in traditional typologies of involvement, from political discussion and persuasion to fundraising and grassroots organizing, can now take place through interactive media. Findings for the study uphold the tenets of the media participation hypothesis while pointing to important political limitations of media participation effects.

## **Innovative Sampling Strategies**

### **Redesigning National School Surveys: Coverage Improvement Using Multiple Datasets**

Alice Roberts, *ICF International*

William Robb, *ICF International*

Kate Flint, *ICF International*

Ronaldo Iachan, *ICF International*

This paper discusses the redesign of the sampling frame development for two national school surveys. Historically, the sampling frames for these surveys have been developed from files acquired from a commercial vendor. Vendor files are used as they contain up-to-date contact information for districts and schools, facilitating recruitment efforts in support of high response rates. The vendor-provided data set incorporates several sets of variables sourced from the National Center for Education Statistics' (NCES) Common Core Data (CCD) and Private School Survey (PSS).

This paper explores the construction of an augmented frame, built by combining the vendor list with NCES files directly, using schools listed on the CCD for public schools and the PSS for non-public schools. A particular challenge is assembling an unduplicated frame for non-public schools in the absence of a unique identifier common across datasets. In addition, with multiple sources for both districts and schools, identifying a unique set of district-school relationships is difficult.

We provide an assessment of eligibility rates from each data source and assess coverage improvements by grade, school type and geography. We also consider the costs and benefits of increasing school coverage with relatively minor increases in the number of students represented by additional schools. For studies with minimum student enrollment thresholds for inclusion, the benefits of the expanded coverage provided by combining multiple datasets may be minimal.

## **Using GIS to Compare Response Patterns for Two Web-Only Recruitment Methods: What Are the Implications for Coverage?**

Lee Florio, *NORC at the University of Chicago*

Michael Stern, *NORC at the University of Chicago*

Ned English, *NORC at the University of Chicago*

One of the primary challenges to Web-only survey use in the general population is the lack of a reliable probability-based recruitment method. Email addresses are infrequently linked to physical addresses, and in cases where such matched lists exist, little is known about their coverage properties. Using data from a Web-only health survey conducted in the state of Illinois, our paper uses Geographic Information Systems (GIS) and geospatial modeling to analyze the potential for spatial bias in response rates while comparing completes in an ABS design recruited via email with those recruited using a URL printed in an advance letter. Given the diverse landscape of Illinois, a Self-organizing map (SOM) is implemented to categorize tracts by salient demographic characteristics as well as infrastructural characteristics like urbanicity, housing unit density and broadband internet penetration. We then examine the distribution of completes by recruitment mode among SOM categories and measure spatial autocorrelation of completes across a hexagonal tessellation using Moran's I. In so doing we explore the kinds of places that tend to have higher responses in web surveys, with a focus on recruitment method. The results from this work inform survey research as to potential geographic coverage issues not previously well understood and allow for tailored and adaptive designs reflecting from an overlooked organizing principle: physical location of respondents.

## **A Spatial Approach to Mitigate Measurement Error from Undercoverage Estimates, with Insights to Improve Address List Representativeness**

Michael Tzen, *U.S. Census Bureau*

In many sample-based surveys, a preliminary and fundamental process is the development of a sampling frame that represents the target population. As many surveys are now creating frames from "address lists," studying a list's features through error sources is vital. Numerous research projects have sought to measure undercoverage of address frames; however, we argue that estimates of undercoverage that incorporate human enumeration are further subject to measurement error. Our study of coverage assumes sampling error is nonexistent (as our data is essentially a census) while mitigating measurement error by a natural way of "smoothing." The U.S. Census Bureau's preferred address list is the Master Address File (MAF). Since addresses approximate geographic locations of housing units, the spatial phenomenon of study is what we will term the MAF "Omission Disease." The terminology draws parallels to "Disease Mapping" applications. We consider a chosen fixed region of study, the state of West Virginia (WV), where the population is the housing units (inhabitants) of WV. We are interested in the inhabitants omitted from the MAF (disease). The model we use is a Generalized Linear Mixed Model with a spatially lagged effect. We present an example where we map "hotspots" and rank these "hot" areas. The rankings are a way to prioritize areas based on the risk of housing unit omissions. This has direct ramifications on operational goals such as redistributing resources for costly field listing procedures. Alternatively, our framework allows for a complementary interpretation of omission associated quantities (e.g. Ruralness, Post Office Completeness, Tenure Status, etc.) previously studied by researchers. With these suggested quantities, our approach also accounts for spatial association. This can be more natural and powerful than alternative methods that ignore "Tobler's Law"; everything is related, but things closer together can be more related than things further apart.

## **A Northern Perspective – Investigating the Application of the Canadian Targeted Address Canvassing Methodology in the United States**

Kathleen Kephart, *U.S. Census Bureau*

Justin M. Ward, *U.S. Census Bureau*

Kevin M. Shaw, *U.S. Census Bureau*

Past research has found that the coverage of the United States Postal Service Delivery Sequence File is sufficient for many areas of the country, but does require supplemental field listing to achieve adequate coverage for some areas (Iannacchione et al 2007). During their 2006 Census, Statistics Canada made a similar observation when they performed a national address canvass. In response, they implemented a targeted listing operation for the 2011 Canadian Census. The Canadian methodology utilized a scoring system to determine which Listing Units required supplemental field listing. As a control, they included some Listing Units that were predicted to have high coverage rates for listing. They found that areas with high coverage were more likely to contain new addresses if they bordered targeted Listing Units (McClellan 2012). Census Bureau researchers are also exploring ways to predict areas that require field listing to augment the address frame prior to the census, and areas where the address frame may be updated with other sources (Tomaszewski 2013, Boies & Shaw 2013). This research explores how the Canadian targeted address canvassing methodology might be applied in a census operation conducted in the United States. We will investigate this using a micro simulation conducted using data available both before and after the 2010 Census Address Canvassing operation. The performance of this methodology will be compared to existing targeted address canvassing models.

## **Differences in Coverage and Nonresponse When Using a Full Household Enumeration Screener vs. a Child-Only Screener in a 2013 National Mail Survey**

Danielle Battle, *American Institutes for Research*

Stacey Bielick, *American Institutes for Research*

Evaluating strategies for within household screening in household mail surveys is increasingly important as address-based mail surveys are used to replace RDD household telephone surveys. This study evaluates whether or not it is possible to effectively enumerate the entire household versus only target populations without affecting response rates and coverage of a particular sub-group. In household studies that seek to sample household members with certain characteristics, a specific household member must be identified and it is essential to enumerate all eligible individuals before selecting the sampled individual. However, researchers must balance the scope of the enumeration (coverage) against enumeration burden, which might affect response rates. Another potential issue is increased likelihood of measurement error in a more complex household roster. The National Center for Education Statistics (NCES) conducted a household enumeration experiment in 2013 to examine these potential issues.

In the experiment, a random sample of 500 addresses received a screener questionnaire that enumerated the entire household; another random 500 addresses received a screener that enumerated only children and youth ages 3 to 20. This child-only screener had previously been used by NCES in a 2012 survey. Expanding it to include all household members would allow NCES to use the sampled household for surveys other than child-focused surveys. We will compare response rates achieved in the two experimental groups and to each other and to response rates achieved in the 2012 survey to determine if there is any negative effect on overall response of requesting information about all persons in the household. We will also look at coverage rates of children between the two groups and the 2012 survey by comparing the

number, age, and grade level of the children reported in each screener, and with population estimates.

**Saturday, May 17**  
**3:30 p.m. – 5:00 p.m.**  
**Concurrent Session J**

## **Interviewing by Texting**

### **May We Text You A Question? An Experiment Using Text Messages For Survey Research**

Kyley McGeeney, *Gallup, Inc.*

Jenny Marlar, *Gallup, Inc.*

Manas Chattopadhyay, *Gallup, Inc.*

Peytchev and Neely (2013), in their article “RDD Telephone Surveys: Toward a Single-Frame Cell-Phone Design,” suggest that in the future, nationally representative surveys in the US could use only a cellphone frame. However the use of live interviewer to call respondents is becoming increasingly expensive and response rates are decreasing, especially among certain demographics such a youth. Reaching respondents on their cell phones using text messages may be an effective alternative strategy.

To explore these alternatives, Gallup conducted an experiment comparing three data collection modes: telephone with live interviewer, text messages and text message with an invite and URL link to a mobile-optimized web survey. Gallup also compared two length treatments. The sample source for this study were cell phone respondents from the Gallup Daily Tracking survey who explicitly agreed to be recontacted via text message on their cell phone. The response rates of all treatment groups were analyzed to determine the feasibility of text message as a mode. . Results will also be used to suggest best-practices for text message surveys.

### **Effort and Sensitivity Effects in Mobile Text Messaging Interviews**

Michael F. Schober, *New School for Social Research*

Frederick G. Conrad, *University of Michigan*

Huiying Yan, *University of Michigan*

Matthieu G. Sauvage-Mar, *GeoPoll*

Text messaging (SMS) is increasingly used to collect mobile responses to survey questions in market research, political polling and social measurement. Although this makes sense given the widespread adoption of cell phones (especially in regions without well developed landline infrastructure for surveys), little is known about how properties of texting affect survey responding. In the study reported here, 2472 respondents from the GeoPoll frame in Tunisia texted answers to 12 closed-form survey questions, texted to them in French by an automated system, in an experiment designed to test two hypotheses. (1) An effort hypothesis is that text survey respondents will be more likely to choose responses that require fewer keystrokes, independent of the content of answers. To test this, we compared distributions of answers for respondents required to key in full responses to distributions for respondents required to answer with a single character (i.e. “1,” “2,” “3,” or “4”). We also examined the effect of character length of an option, to see if respondents are drawn to shorter-character response options in the multi-character condition. (2) A sensitivity hypothesis is that text survey respondents may be more likely to select the more sensitive response options when they are single-character (that is, with

arbitrary rather than meaningful and thus potentially value-laden labels) than when they must articulate (type) the full answer. Results demonstrate overwhelming support for the effort hypothesis: respondents in the multi-character condition were more likely to select the shortest response option than respondents in the single-character condition. The sensitivity hypothesis was supported for one question: respondents were more likely to report having paid a bribe with a single-character response than by typing out “oui” (“yes”), which is no shorter than “non.” Of course, we don’t know the true values, but these results begin to clarify principles for designing text surveys.

### **Interviewing by Texting: Costs, Efficiency and Data Quality**

Frederick G. Conrad, *University of Michigan*

Michael F. Schober, *New School for Social Research*

Christopher Antoun, *University of Michigan*

Andrew Hupp, *University of Michigan*

As people’s communicative habits change with the rise of mobile devices, collecting survey data via SMS text messaging is an increasingly attractive alternative to traditional telephone interviewing. But the properties of text interviews—from the quality of the data to their operational costs—are not yet well understood. Here we report on interviewer efficiency and production costs in a study of 648 voice and text interviews, half administered by interviewers at the University of Michigan Survey Research Center and half administered by automated systems developed for the project. Respondents, who had opted into participating in a survey on their iPhone, were randomly assigned to answer the same 32 questions via either voice or text, with a human or automated interviewer; as in voice, interviewers administered one text interview at a time. Text interviews achieved a higher response rate than voice interviews, with fewer contact attempts and less total field time per interview. They also took much longer than voice interviews, as there are much longer intervals between turns in text. In comparison with automated text interviews, human text interviewing improved completion rates, while automated interviews allowed efficient contact attempts and of course had no interviewing costs. Text interviewing would be yet more efficient if interviewers administered more than one text interview at a time, which interviewers reported would be desirable. These findings add to an emerging picture of the considerations for choosing to administer text vs. voice interviews. We argue that the choice depends on which priorities are highest: data quality (here, precision of responses and disclosure), respondent satisfaction, data management, respondent burden, breakoff rates, and costs for interviewer time vs. software development for automated interviewers. We propose that new measures of cost and efficiency beyond those developed for telephone interviews are needed for less synchronous modes like text interviewing.

### **Text Message Follow-Up for Cell-phone Respondents to a National Survey**

Ashley Amaya, *NORC at University of Chicago*

Jennifer Vanicek, *NORC at University of Chicago*

Michael Stern, *NORC at University of Chicago*

Carla Black, *Centers for Disease Control and Prevention*

Cindi Knighton, *Centers for Disease Control and Prevention*

Can text messages be an effective refusal conversion technique in a public health survey? In surveys that have sampled addresses, researchers may opt to mail a refusal conversion letter and/or an incentive to explain the survey, offer the benefits, and encourage participation. The same protocol cannot be used for a cell RDD sample since an address is usually not available. In Quarter 3 of the 2013 National Immunization Survey (NIS), CDC and NORC launched an



experiment to test the effectiveness of text messages as a refusal conversion technique. Text messages were sent to households with NIS age-eligible children that had broken off during the telephone interview but consented to receive a follow up via SMS/text message. The experiment tested multiple conditions, including an invitation to complete the survey on the internet (rather than by telephone) as well as the effect of a promised \$10 incentive upon completion of the interview. Results of this experiment indicated that respondents were not responsive to text message follow up; however, sending the messages did not have a negative impact on refusal conversion rates. This paper will briefly describe the process used to identify respondents and send text messages. It then explores the effects of text message follow up on interview completion both by telephone and web, the effects of including a post-paid incentive offer in the text message, and effects on refusal rates. Overall, text message follow-up was observed to have very little effect, positive or negative, on respondent behavior. This is likely due to the relative novelty of receiving text messages from survey as well as governmental organizations, rather than personal or more traditional individually orientated businesses (e.g., one's friends or personal bank, respectively), sources. Respondents may simply not expect or react to text message prompting.

## **Methodological Briefs: Survey Research in Educational Settings**

### **Can a Survey of U.S. High Schools be Replaced or Reduced Through Web Searches? The Successes and Complications of an Experimental Strategy**

Casey Langer Tesfaye, *American Institute of Physics*

Susan C. White, *American Institute of Physics*

Because physics is such a small field, data about physics in high schools is not readily available from traditional sources. As a result, in 1986 the American Institute of Physics began a quadrennial survey of high school physics. This survey has two phases. In the first phase, we contact a random sample of the public and private schools in the U.S. to find out whether the school offers physics and to collect teacher contact information from those that do. In the second phase, we contact the teachers we have identified directly with more detailed questions. This survey effort has traditionally included a combination of notification letters, paper, phone and web surveys for schools and then for the teachers. In the past our school survey has enjoyed a consistently high coverage rate, with only a handful of refusals to participate and a response rate over 95%. However, we have encountered new challenges over the past few survey rounds, including more refusals to participate at both the district and school levels, due in part to a rapid increase in school surveys from a number of other sources. In the 2012-2013 academic year, we experimented with a new strategy, beginning our effort first with a series of web searches to see if the information we traditionally collect from the school survey was available online. The goal of this strategy was to eliminate or dramatically reduce the size of the school survey. We will explore the viability, successes, unique complications and lessons learned from this effort.

## **Methods for Collecting Physical Activity Data in Urban School Settings**

Brittany A. Vas, *Mathematica Policy Research*

William Reeves, Jr., *Mathematica Policy Research*

Martha Bleeker, *Mathematica Policy Research*

Nicholas Beyler, *Mathematica Policy Research*

Kellie Borradaile, *Mathematica Policy Research*

Given the rise of school-based interventions aimed at helping students become more active, there is a critical need for development of accurate methods for assessing physical activity in school settings. The present study used a multi-method approach to comprehensively capture physical activity outcomes during recess, while balancing competing demands to increase efficiency, decrease burden on schools, and maintain the integrity of a rigorous, randomized control study. In particular, this presentation will draw on experiences using two physical activity measures at 29 schools as part of an evaluation of Playworks, a program that places full-time coaches in low-income schools to provide opportunities for organized play during recess. ActiGraph GT3X accelerometer devices were used to collect student-level data related to the intensity of physical activity and number of steps taken during recess. The System for Observing Play and Leisure Activity in Youth (SOPLAY), in contrast, provided data on contextual, school-level activity patterns and documented the types of activities in which students were engaged through systematic observation of students in zones of the recess play yard. The combination of measures provided unique insights into factors that may have influenced the impact of Playworks on students' physical activity.

This presentation will discuss differences in findings between the two measures and address challenges associated with both. For example, the accelerometers provided important student-level precision that could not be obtained through direct observation, but also required us to obtain active parental consent for a new technology with which many parents, students, and teachers were not familiar. The SOPLAY observation tool, in contrast, provided detailed, contextual information to help explain the accelerometer findings, but required creative adjustments and extra planning, including preliminary trips to each school, so that the tool could be used to accurately measure students' activity in small, high-density recess spaces in urban school settings.

## **Effects of a Longitudinal Measurement Burst Design on the Retention of Students Graduating from High School**

Jamie Griffin, *University of Michigan*

Megan E. Patrick, *University of Michigan*

Carlos Macuada, *University of Michigan*

Research examining the transition to adulthood is often restricted to college student samples because sampling frames are readily accessible. The little that is known about young adults not in college comes from a few well-known longitudinal studies (e.g., Bureau of Labor Statistics, 2012; Johnston et al., 2013); however, those studies do not collect daily-level data. Because these larger studies have documented that various paths to adulthood are characterized by different patterns of risk behaviors (Schulenberg et al., 2005), it is important that studies of young adults consider young adults on all paths.

The purpose of the current study was to determine how effectively a longitudinal measurement burst design (Sliwinski, 2008) assessing daily-level associations between substance use, predictors, and consequences retained young adults who did and did not attend college. High

school seniors (N=318) from three Midwestern schools completed an in-school baseline survey in the spring of 2012. Respondents were then randomized into an intensive measurement (IM) or control group. Four and eight months after baseline, young adults in the IM group received a 30-minute follow-up web survey, followed by 14 days of web-based daily surveys. Twelve months after baseline, young adults in both groups were invited to participate in a final 30-minute follow-up web survey.

Notably, we found that the IM design did a better job of retaining young adults who indicated that they did not want to attend a four-year college or graduate/professional school than did the control design. The discussion will also include the demographic representativeness of Wave 3 respondents in both measurement groups (e.g., disproportionately more females retained by the IM design) and differences in substance use by measurement group (e.g., IM group reported less binge drinking than control group). Together, these findings suggest that success of the IM design depends on the sample characteristic of interest.

### **How to Ask for Parental Permission to Interview a Teenager: A Telephone Experiment in the California Health Interview Survey (CHIS)**

Matt Jans, *UCLA Center for Health Policy Research*

David Grant, *UCLA Center for Health Policy Research*

Royce Park, *UCLA Center for Health Policy Research*

Sherman Edwards, *Westat*

One challenge of interviewing adolescents is obtaining parental consent. This can be particularly difficult in telephone surveys where parents have no personal connection with the researchers. The California Health Interview Survey (CHIS) has experienced a decline in adolescent interview yield over the past 10 years due primarily to declining parental permission. In an attempt to counter this trend, focus groups were conducted with parents to get a better understanding of their concerns. From that feedback, an experiment was designed, which changed the way permission was requested. The randomized experiment manipulated a) the verbatim consent script read to parents, with the new version using plainer and friendlier language without changing any of the substantive information about the interview or research risks, and b) instructions to interviewers to tailor responses to parents' questions, in essence motivating them to participate if they express ambiguity or resistance. Initial results show that the new permission script significantly increases the rate at which parents grant permission to interview the adolescent. Encouragement to tailor response to parents' questions seems to have no additional positive effect and may negatively affect permission rates when combined with the new consent script. The overall effects seem to be driven by differences in Spanish interviews but not English interviews, suggesting that either translation or cultural expectations regarding parental requests for adolescent participation may influence parental consent rates. The findings contribute to our understanding of methods to include minors in population-based surveys, and language differences in survey consent and participation.

### **Understanding the Role of Biases and Prejudice in Assessments of Students and Preferences for Educational Policy**

Rachel L. Moskowitz, *Northwestern University*

This laboratory experiment explores how race, class, and school quality affect both the assessments of individual students' noncognitive skills as well as the willingness to support the closing of neighborhood schools. This project attempts to tie education psychology research with political science scholarship on the role of prejudice on policy attitudes and notions of

equality. Participants are provided with randomized information about the student's socio-economic class, school quality, and visual cues about his race. The study then looks at differences between the beliefs about students in order to differentiate between perceptions of the public about the motivations of low-income and minority students. Additionally, it explores how these beliefs about students and their demographic characteristics then impact support or opposition to school closings. This also includes attitudes about dimensions of equality and community in the specific neighborhood. While race is predicted to be the most powerful explanatory factor for the public's beliefs about issues related to individual students' motivations and mindsets, the quality of the school the student attends should greatly impact attitudes about school closings and equality.

### **Truant from the Data: Nonresponse and School-Based Survey Administrations**

Daniel G. Harwell, *American Institutes for Research*

Sandra Eyster, *American Institutes for Research*

Samantha Neiman, *American Institutes for Research*

Russell C. Brown, *Cleveland Metropolitan School District*

As response rates to surveys have continued their downward trend, survey researchers have grown increasingly concerned about the potential for nonresponse bias; however, as Groves (2006) discusses, response rates are actually a poor predictor of nonresponse bias. In fact, nonresponse is a major cause for concern any time the topic of interest is directly related to response propensity.

This study examines data from the Conditions for Learning (CFL) Survey, which is administered online to all students in the Cleveland Metropolitan School District and measures students' perceptions of school climate in four domains: school safety, support, challenge, and social and emotional learning. The survey achieves an approximately 80 percent response rate, thanks in part to the school-based administration procedures. School-based student surveys are becoming increasingly more common, and are often administered by staff who do not have expertise in issues of survey administration, nonresponse bias, and survey procedures that emphasize nonresponse follow-up.

Response propensity may be related to the measures of interest in student perception surveys. For example, students who feel unsafe in school or who are performing poorly academically may be less likely to attend school they and therefore complete the survey, leading to a potential for nonresponse bias. Using administrative records regarding student characteristics, including academic performance and attendance, this study will examine the potential for nonresponse bias in these data and discuss the implications for the use of school-based student perception surveys.

### **Increasing Assessment Coverage: Extending a School-Based Mathematics Assessment to Include Out-of-School Administration**

Randolph Ottem, *RTI International*

Debbie Herget, *RTI International*

Amy Kowalski, *RTI International*

Education studies that include a school-based assessment often collect assessment data only from students able to participate in an in-school session. For example, the National Education Longitudinal Study of 1988 (NELS:88) and the Education Longitudinal Study of 2002 (ELS:2002) relied exclusively on in-school make-up sessions to collect data from nonresponding

students, which excluded students who left the base-year school. By only assessing students in schools, these studies had gaps in assessment coverage due to not following up with students who were unable to participate in-school. Other studies, however, made a more costly decision to administer make-up assessments in-person. The High School and Beyond study, for example, used field interviewers to administer the assessment in locations such as a library or church, or occasionally in the student's home. The first follow-up to the High School Longitudinal Study of 2009 (HSL:09) attempted to collect self-administered mathematics assessment data from all student respondents, regardless of whether the student participated in- or out-of-school. This paper analyzes the increased coverage provided by collecting computerized, self-administered assessments outside the school setting, by mode of completion (i.e., in-school, self-administered Web, or in-person), high school enrollment status, key demographics, and assessment scores attained by students who participated in- and out-of-school. This paper will investigate several research questions: (1) what is the increase of assessment coverage by offering the assessment outside of the school; (2) what is the difference in assessment response among students who completed the questionnaire via Web, telephone, or in-person with an interviewer; (3) are the demographics of students who complete the assessment outside of school significantly different than those who complete the assessment in the school; and (4) are assessment scores significantly different for students who participate in-school than those who participate out-of-school?

## Public Opinion Across the Globe

### **The Welfare State and Attitudes Toward Inequality and Redistribution: Data From 46 Nations and 65,000 Respondents**

Jonathan Kelley, *University of Nevada-Reno, International Survey Center*

Mariah Evans, *University of Nevada-Reno*

Nate Breznau, *Bremen International Graduate School of Social Sciences*

Income inequality is a central focus of political struggle in the modern world with a striking difference between hands-off free-market systems, mainly in English-speaking lands, which largely limit government action to providing "rules of the game", a stable legal framework, a market-oriented economy, and a limited social safety net; for them inequality is a normal "side effect" of the market economy. At the other extreme are welfare states, many emerging in Northern Europe after World War II, whose governments champion extensive intervention in the economy: corporatist labor markets, income redistribution, and generous social welfare. In this paper we ask how their citizens have come to view these different political systems. Data are from the World Inequality Study including 112 surveys in 46 countries with over 65,000 respondents, mostly from the ISSP's four Inequality modules, 1987 to 2009. The weakest theory is that government policy is irrelevant (Marx, classical economics, other self-interest theories); the strongest that government is everything (dominant ideology, system justification). Our results reject both, instead suggesting institutional anchoring theories, with the core attraction of systems strongly valued but secondary attractions accepted mainly because of that. The core features of the welfare state could be helping the poor, or reducing inequality, or furthering redistributive beliefs – all prominent themes in its rhetoric. But those matter little. Instead the core attraction appears to be envy (or possibly resources for government): modest pay for high status occupations, much less than in market-oriented societies. In the strongest welfare states (Austria, Sweden) people favor paying top occupations 30% or 40% less than would be found legitimate in the most market-oriented societies (US, Chile). But no increase in the pay of low

status occupations. Thus welfare states' core attraction is cutting down the rich but not raising up the poor.

### **The Millennium Development Goals: Using Public Opinion to Help Policymakers Measure Progress and Prioritize Goals**

Neli Esipova, *Gallup, Inc.*

Zach Bikus, *Gallup, Inc.*

Dato Tsabutashvili, *Gallup, Inc.*

Since their inception in 2000, the Millennium Development Goals (MDGs) have provided a framework for improving the lives of the world's poorest individuals. Apart from national governments, data is obtained through surveys carried out every 3-5 years by a wide variety of organizations and agencies, making direct comparisons across countries difficult. Using Gallup World Poll data, this paper will explore the possibility of annually tracking the progress using data with a single methodology worldwide. This data is drawn from nationally representative samples consisting of at least 1,000 interviews in each country, with more than 1.2 million total respondents since 2006.

We will analyze the relationship between evaluative wellbeing (which is strongly correlated with a country's development level) through respondents' evaluation of their own lives and other variables that can be used as a measure of progress towards goals such as good governance, poverty, security, and equality for women. We will examine the possible effect of existing and proposed post-2015 Millennium Development Goals on wellbeing using regression analysis which will allow us to make predictions on how achieving certain MDG goals will improve people's wellbeing in countries of different developmental levels. Individual countries may see the wellbeing of their citizens positively affected at varying levels of magnitude while achieving each of the MDG goals. While one country may see a large improvement in wellbeing from improved security, others may find that wellbeing increases most through greater food security. As such, our research will help policymakers to track progress annually, measure the overall benefit of reaching the goals, and to prioritize goals that will have the greatest impact to their specific country.

### **Outside Looking In: An Examination of the Kaleidoscopic Nature of International Public Opinion of the United States During the Bush and Obama Presidencies**

Natalie Manayeva, *University of Tennessee*

Alexandra Brewer, *University of Tennessee*

Michael Fitzgerald, *University of Tennessee*

Public opinion polls from around the world demonstrate that during the past decade the United States' international image has worsened and approval for U.S. actions has declined (Fitzpatrick, Kendrick & Fullerton, 2011). This negative trend has changed in 2008, when Barack Obama was elected President, and the favorable attitude towards United States have spiked worldwide. However, by the end of the first year of Obama's second term in the White House those positive tendencies are vanishing around the world, and the leadership of the U.S. is drastically losing popularity.

Strengthening of the anti-American sentiments across the globe presents a variety of negative consequences for the United States. Negative attitudes towards the country may result in economic and political losses, and could even cause serious international conflicts (Revel,

2003). Resolving the problem of rising anti-Americanism and dealing with its negative outcomes requires understanding of the phenomenon, its origins and mechanics.

The origins of anti-American feelings and attitudes have been studied by scholars in various disciplines and approaches. Katzenstein and Keohane (2007) distinguished six types of anti-Americanism, which varied in causes and features. Other scholars (Crockatt, 2003; Meunier, 2005) identified a variety of historical, cultural, religious, and economic reasons for anti-American attitudes.

This study is designed to explore the multi-faceted nature of international public opinion towards the United States during the Bush and Obama presidencies and to provide possible explanations for the fluctuations of global attitudes by analyzing domestic and non-domestic factors. Data from Gallup polls on the attitudes towards the United States is analyzed in this study. The ultimate goal of this research is to expand understanding of the linkage between the international public opinion of the United States and the factors of U.S. domestic and foreign policy.

### **Civic Engagement Among First-Generation Migrants Worldwide**

Julie Ray, *Gallup, Inc.*

Neli Esipova, *Gallup, Inc.*

Anita Pugliese, *Gallup, Inc.*

The well-being of the world's migrants is emerging as a vital part of post-2015 development discussions. The problem with informing these discussions from a public opinion research perspective is that few studies exist on the lives of international migrants as a subgroup and none has ever been carried out on a global scale. From nearly 500,000 interviews conducted as part of its World Poll in 150 countries between 2009 and 2011, Gallup identified nearly 25,000 first-generation migrants. This large sample enabled Gallup to analyze this cohort's well-being from multiple angles, comparing it with the native-born in their adopted countries, examining how length of stay affects it, and using models to estimate potential gains and losses from migration.

Studying migrants' civic engagement is important because of the strong, positive relationships Gallup finds between it and multiple aspects of their well-being. This paper measures migrants' commitment to the communities they live in by assessing their inclination to donate money to charity and volunteer their time to an organization, and explores how this commitment changes over time.

### **The Role of Door-to-Door-Campaigning in the 2013 German General Election Campaign**

Nicole N. Podschuweit, *University of Mainz*

Thomas Roessing, *University of Mainz*

In the run-up to the 2013 German Bundestag election, there was one campaign tool many national media reported on: Surprisingly, it was no matter of the Internet, but face-to-face conversation at people's homes. According to media coverage, door-to-door-campaigning was one of the most promising campaign tools and, in contrast to several online gimmicks, had the potential to affect voting decisions. The present paper addresses the question if door-to-door-campaigning was as important for the outcome of the election as the mass media suggested. To answer this question, we prepared an online survey among campaign aids of the Social

Democrats and the Christian Democrats. The survey is scheduled to go into the field during the remaining weeks of 2013. The respondents will be addressed over parties' official communication channels. According to media reports, the Social Democrats delegated more than 14,000 campaign aids for door-to-door-campaigning. As the election district candidates made up the majority of the door-to-door-campaigners of the Christian Democrats, the population in this case is about 300. The questionnaire contains questions on the amount of persuasion attempts, the duration of achieved talks as well as on characteristics of the campaign aids and of the voters. Furthermore, we will ask questions on the reactions to campaign aids' approaches and campaign aids' motivation for supporting the door-to-door-campaigns. This study will allow drawing conclusions on the relevance of interpersonal communication as a campaign tool in Germany. Beyond that, the results of the survey may necessitate a rethinking of the relevance of interpersonal communication in the media age.

## **Topics in Weighting and Adjustment**

### **Using a Sample Weight Adjustment to Account for Poverty-Related Survey Error in the California Health Information Survey**

Tara L. Becker, *UCLA Center for Health Policy Research*

YuChing Yang, *UCLA Center for Health Policy Research*

Hongjian Yu, *UCLA Center for Health Policy Research*

Yueuan Wang, *UCLA Center for Health Policy Research*

With the implementation of the Affordable Care Act well-underway, the importance of collecting high quality income and poverty data in order to estimate the number of families and individuals eligible for state and federal health insurance programs has increased for health-related surveys. One ongoing source of concern for the California Health Information Survey (CHIS) is that the distributions of income and poverty in California diverge from the well-established measure of poverty estimated by the U.S. Census Bureau using the Current Population Survey (CPS). CHIS consistently estimates that a higher proportion of Californians are in either the lowest (below the federal poverty line (FPL)) or the highest (400% or more of FPL) income groups, relative to the CPS. This potentially overstates the population eligible for state health insurance programs, such as Medi-Cal, and understates those eligible for federal health insurance subsidies on the state's health insurance exchange, Covered California. One way to bring CHIS estimates in line with the CPS is to treat this disparity as the result of survey error and create a new set of CPS-adjusted weights. These weights add a new raking dimension that is based on the distribution of poverty from the CPS. When the reweighted distributions of income and poverty status are compared, the CHIS sample more closely approximates that of the CPS. Using CPS-adjusted weights also reduces the proportion of Californians eligible for Medi-Cal and increases the proportion with health insurance coverage through their employer. We explore the effects of using this new set of weights on these and a variety of related measures in CHIS. We discuss the benefits and disadvantages of using these new weights and provide researchers with some guidelines on when they should consider using them rather than the standard CHIS sampling weights.



## **Evaluation of Alternative Weighting Approaches to Reduce Nonresponse Bias**

Andy Peytchev, *RTI International*

Kumar Rao, *The Nielsen Company*

Michael W. Link, *The Nielsen Company*

Ceril Shagrin, *Univision*

With declining response rates, surveys increasingly rely on weighting adjustments to correct for potential nonresponse bias. The resulting increased need to improve survey weights faces two key challenges. First, additional auxiliary data are needed to augment the models used to estimate the weights. Depending on the properties of these auxiliary data, nonresponse bias can be reduced, left unchanged, or even increased (e.g., Lin and Schaeffer, 1995). Thus, the second challenge is to be able to evaluate alternative weighting approaches, when the assumption of “change in estimates means less bias” may not hold. Ideally, data need to be collected from nonrespondents to provide direct estimates of nonresponse bias.

The Nielsen TV Diary, a national probability-based survey, has experienced the ubiquitous trend in declining response rates. In 2012, a nonresponse bias study was conducted in three geographic areas. An abbreviated instrument to evaluate nonresponse bias was mailed to Diary nonrespondents and respondents. A random sample of remaining nonrespondents was selected for in-person interviewing, increasing the three areas' response rates to 58%, 68%, and 73%. We then computed several alternative weighting adjustments to produce estimates for the Diary respondents, and compared them to the estimates based on the Diary respondents and nonrespondents.

There are two general ways to improve weighting adjustments: expand the types of auxiliary data and change how the auxiliary data are modeled. We evaluated four sets of weights: the current approach based on poststratification, adding sample-based nonresponse adjustment based on matched telephone number (Link and Lai, 2011), adding census geocoded data (e.g., Biemer and Peytchev, 2013), and adding interaction effects in poststratification. We describe the ability of each approach to reduce nonresponse bias and suggest avenues for future research. We also note the underutilization of two stage sampling for nonresponse as a cost-effective method to produce direct estimates of nonresponse bias.

## **Effects of Differing Weights on Regression Coefficients**

Hee-Choon Shin, *National Center for Health Statistics*

Jibum Kim, *Sungkyunkwan University*

A set of unweighted normal equations for a least squares solution assumes that the response variable of each equation is equally reliable and should be treated equally. When there is a reason to expect higher reliability in the response variable in some equations, we use weighted least squares (WLS) to give more weight to those equations. For an analysis of survey data, sampling weights are typically used for unbiased and efficient estimates. There might be reasons for deviating from these weights – e.g., heteroscedasticity or extreme weights. Different weights can yield different point and interval estimates of the coefficients, affecting the interpretation of results. We consider the impact of differing functional forms of weights on the WLS solutions using data from the 2009-2010 National Health and Nutrition Examination Survey (NHANES), a periodic survey conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC). The following five weightings are considered: (1) Constant; (2) Sampling weights; (3) Positive square root of the sampling weights; (4) Inverse of the estimated variance; and (5) Square root of the inverse of the

estimated variance. We model body weights (Kg) as a function of heights and other explanatory variables including sex and race/ethnicity, and demonstrate the effects of using different functional forms of weights on the regression coefficients.

### **Nonresponse Bias in National Health Surveys — Kingdom of Saudi Arabia, 2013**

Charbel El Bcheraoui, *Institute for Health Metrics and Evaluation*

Margaret G. Robinson, *Institute for Health Metrics and Evaluation*

Paria Naghavi, *Institute for Health Metrics and Evaluation*

Marwa Tuffaha, *Institute for Health Metrics and Evaluation*

Sarah L. Mikhitarian, *Institute for Health Metrics and Evaluation*

Farah Daoud, *Institute for Health Metrics and Evaluation*

Ali H. Mokdad, *Institute for Health Metrics and Evaluation*

The 2013 Saudi Health Interview Survey was designed to provide regional and national estimates of non-communicable diseases and their risk factors. Households were selected through a multi stage sampling strategy, then an adult was randomly selected to complete a computer assisted personal interview. Blood pressure, weight and height were measured for all respondents who were invited to visit medical clinics for further examination, including blood chemistry for lipids, diabetes, and vitamin D. A total of 12,000 households were selected, 11,730 head of the households completed a roster, and 10,827 selected adults completed the survey (92.3% participation rate). However, only 4,730 (43.7%) of these visited the clinic for the medical exams. Women, obese, and individuals previously diagnosed with pre-diabetes were more likely than their counterparts to visit the clinic for further examinations. To account for self-selection bias in completing the clinic-based examination, we applied a post stratification adjustment using socio-demographic characteristics, health behavior, previously diagnosed non-communicable diseases, and anthropometric measurements of respondents from the household survey. Without our post stratification methodology, this self-selection bias would have overestimated the real prevalence of non-communicable diseases.

Our study showed the impact of self-selection bias of individuals who participated in a clinic-based examination in a national survey. Our household data enabled us to implement a post stratification adjustment for this selection bias. Indeed, in such national survey the burden of non-communicable disease would have been over estimated.

### **Experiments in Weight Trimming**

Benjamin Phillips, *Abt SRBI*

Stanislav Kolenikov, *Abt SRBI*

Survey weights incorporate multiple adjustment factors including probability of selection, frame integration, within-household selection and other multiplicity factors, nonresponse, and raking. Nearly all of these procedures introduce additional variance to the weights, increasing design effects. Trimming, reducing especially large weights to a maximum (and possibly minimum) threshold, is a commonly used remedy for excess variance. Various heuristics exist for this purpose (e.g., trim above 5x mean weight, median + 6x IQR, 99th percentile). Trimming, however, may introduce bias to the survey estimates, potentially negating whatever gains in precision are made by the reduction of sampling error.

We investigate several approaches for reducing weight variance with respect to evaluating variance and bias, when to trim weights, and using compression as an alternative to trimming. With respect to evaluating trimming points, mean squared error (MSE) balances variance and bias. We consider definitions of MSE: with bias calculated from the difference between the

trimmed and untrimmed weights; with bias calculated from the deviation of the weighted estimates from raking marginals; and a cross-validation rule in which one of the raking margins is removed at a time, the trimming cut point is optimized to minimize the MSE of the survey estimates for the excluded margin, the process is repeated for all raking margins, and the ultimate trimming parameter is obtained by aggregation of the margin-specific trimming cut-off points. We also consider when to trim weights: in the simplest setting, trimmed occurs after raking; alternately, weights can be trimmed during each raking iteration. Finally, we consider considering reducing weights above a threshold by a coefficient rather than reducing to the threshold, preserving differentiation in the size of weights above the threshold.

We evaluate the performance of these rules and approaches in an extensive full factorial simulation with a realistic population, sample design, and nonresponse mechanism.

## **Incentive Impact**

### **Impact of Increasing Incentive Values and Cooperation Rates: Lessons from the 2013 Survey of Consumer Finances**

Micah Sjöblom, *NORC at the University of Chicago*

Cathy Haggerty, *NORC at the University of Chicago*

During the 2013 Survey of Consumer Finances (SCF) data collection, incremental cash incentive offers were used to encourage participation from “late responders” as an enhancement to interviewer gaining cooperation and refusal conversion activities. While the use of additional incentives beyond the base cash incentive has played a role in increasing SCF participation overall, the staged process for incremental incentive offers present opportunities to better understand the interaction between incentives and data collection costs. For this presentation we will look at the relationship between incremental incentives and three different factors: conversion outreach efficiencies, participant cooperation and data quality.

To better understand the role incremental incentives play in conversion outreach efficiencies we will analyze the number and different types of contacts required to complete interviews both before and after incentives increases were offered across two distinctive time periods. We will also explore the relationship between incentive interventions and participant cooperation through analysis of interviewer observations collected in the call records regarding the cooperativeness of the household and the household’s reaction to the increased incentive offer. Finally, to look more closely at the relationship between incentives and data quality, we will use interviewer paradata and questionnaire data quality scoring to build comparisons between late-responders offered at least one increased incentive treatment and early responders who were not presented with an increased incentive.

### **Dollars and Sense: Incentives Testing Supporting Quality and Business**

Lauren A. Walton, *The Nielsen Company*

Kelly Bristol, *The Nielsen Company*

Christine Pierce, *The Nielsen Company*

Tracie Yancey, *The Nielsen Company*

Chrystal McQueen, *The Nielsen Company*

Incentives are often a larger component of a survey organization’s budget. Upfront token cash incentives have repeatedly been shown to be very effective at increasing survey participation.

However, as response rates to most surveys have continued to decline, this can result in sending out large amounts of cash to people who never become survey participants. Contingent incentives, promised and sent after participation, are increasingly being used as alternatives, especially when budget concerning. This paper presents the results of three innovative incentives tests that leverage electronic and contingent incentives. These tests were conducted in 2013 using the Nielsen TV diary studies employing an address-based sampling design to recruit participants using a two-stage mixed-mode design. Study 1 tested mailing a single letter requesting participation in a short online survey with the promise of a \$5 or \$10 Amazon.com gift card. Analysis indicates a much lower response rate to this method, when compared with providing \$2 or \$5 cash in the recruitment letter upfront. Study 2 tested different amounts (\$25 vs. \$30) and different bill denominations (\$30 or \$25 in \$5 bills vs. \$30 in \$10 bill and \$20 bill). Results indicate that cooperation was lower among those who received \$25 vs. \$30 and there was a slight, positive non-significant increase in cooperation when providing the \$30 in five-dollar bills vs. the \$10 and \$20 dollar bills. Study 3 focuses on those who refuse to participate when contacted by phone but are still mailed the paper diary to see if they will complete it. In the envelope, test households receive either \$0 or \$2 with the promise of a large contingent cash incentive (\$50, \$75, or \$100). This paper examines contingent vs. noncontingent incentives, large pay for performance incentives, and electronic incentives, balancing methodological best practices against practical business constraints.

## **The Effects of Electronically-Administered Incentives on Web Survey Response Rates**

Jared Coopersmith, *Mathematica Policy Research*

Lisa K. Vogel, *Mathematica Policy Research*

Kathleen Feeney, *Mathematica Policy Research*

Tim Bruursema, *Mathematica Policy Research*

Providing meaningful incentives demonstrates to respondents that researchers understand the competing demands on their time and value their input. The effects of incentives, particularly when prepaid, are strongly established in survey research literature as effective tools for increasing response. However, effectiveness of incentives on web-administered surveys is less clear, and can be impacted by a number of factors, including incentive type and amount, timing, and mode of survey administration.

This paper seeks to contribute to research on incentives for web-based surveys by examining the effects of the following:

- (1) An additional post-incentive for early completion
- (2) A pre-paid incentive in combination with a post-paid incentive
- (3) Pre-incentives as a nonresponse follow-up strategy

This paper will use data collected during the first wave of a nationally-representative survey of public school principals designed to take place across three waves of data collection. To determine a maximally-effective incentive strategy for subsequent years, we embedded an experiment into the study using electronic gift cards. All sample members were eligible for a standard \$50 post-response incentive, but were also randomly assigned to one of four experimental conditions: (1) an additional \$50 incentive for completing early in the field period, (2) a \$25 incentive pre-paid with the initial survey mailing, (3) a \$25 pre-paid incentive used during nonresponse follow-up, and (4) the standard \$50 post-response incentive, which serves as the control group for this experiment.

Differences in AAPOR4 response rates between the groups will be compared using t tests, correcting for multiple comparisons. We will also estimate logistic regression models to further test the effect of the different incentive strategies on the likelihood of response, controlling for a set of covariates. Results will contribute to the research community's understanding of how different incentive strategies can maximize response.

### **Use of Promised Incentives and Methods to Increase Incentive Saliency**

Douglas Williams, *Westat*

David Cantor, *Westat*

Shannan Catalano, *Bureau of Justice Statistics*

The offer of monetary incentives has become a ubiquitous practice for self-administered surveys. Generally prepaid incentives have been found to be more effective than promised incentives (Church, 1993). Research has not found promised incentives to work on a consistent basis (Edwards, et al., 2005). However, sometimes a promised incentive is the only way to provide an incentive to all sampled participants. When used in conjunction with a pre-paid incentive, a promised incentive also offers the opportunity to provide additional motivation to respond. In this paper, we report on experiments aimed at testing a promised incentive, along with a method to increase the visibility of the offer, as part of an Interactive Voice Response (IVR) survey. One experiment offered the promise of \$10 for completing the IVR survey. Another experiment manipulated the saliency of the survey request. These experiments were implemented in a 2012 mode experiment conducted with the National Crime Victimization Survey (NCVS). With respect to response rates, we found a significant increase in response when a promised incentive was offered. Further, this increase was magnified when the incentive offer was made more salient. This suggests previous null findings for promised incentive may be in part due to lack of awareness by the respondent of the incentive offer. Despite the increase in response, the incentive did not produce a difference in the demographic distributions of the sample. However, we did observe significantly fewer breakoffs in the IVR interview for those respondents offered the incentive. Despite the increase in response rates, the incentive did add somewhat to the cost of the survey.

## **Designing Response Scales to Improve Survey Measurement**

### **Investigating Response Quality in Mobile and Desktop Surveys: A Comparison of Radio Buttons, Visual Analogue Scales and Slider Scales**

Vera Toepoel, *Utrecht University*

Frederik Funke, *University of Mannheim*

Mobile devices have smaller displays, touch screens and different methods of navigation compared to desktop computers. This may limit the amount of information that can be placed on a mobile phone screen and it can also affect how a survey is comprehended and completed.

The most traditional rating scales in Web surveys are made from radio buttons. Radio buttons require quite a lot of space. Only a limited number of response options can be presented simultaneously. Otherwise, respondents have to scroll to see all options which may bias ratings.

Visual Analogue Scales (VAS) are operated by point and click: respondents move the mouse arrow to any position on the line and after clicking the mouse button a marker appears. In contrast, slider scales have a handle visible directly on load of the Web page and ratings are

done by drag and drop. Both scales can either be implemented as discrete or continuous rating scales. A continuous implementation is especially valuable if respondents use mobile devices like smart phone where an efficient use of space is required.

A comparison is needed of radio buttons, VAS, and slider bars to see how they affect usability and data quality on mobile phones compared to regular desktop completion. Finger navigation on mobile phones is less precise than mouse navigation on desktops. This could result in selecting the wrong (not intended) answer option in radio buttons. Slider bars or VAS might be more efficient in selecting the intended response option. We look at response quality indicators, paradata, evaluation of the questionnaire as well as personal characteristics. The usability of question formats is conjectured to be related to the number of scale points. We use an experimental design with question format and 5, 7, 11 and continuous scales. Data are collected in a probability-based panel in the Netherlands.

### **The Impact of Question and Scale Characteristics on Scale Direction Effect**

Lirui He, *University of Michigan*

Ting Yan, *University of Michigan*

Florian Keusch, *University of Michigan*

Saram Han, *University of Michigan*

Rating scales are used extensively in surveys and polls. Survey literature has demonstrated that many design features of rating scales affect how survey respondents process the scale and use the scale to construct their responses such as the numerical values assigned to a rating scale, the shape of a scale, the spacing of the response options, and even the color of the scale points. One design feature that hasn't gained as much attention is the direction of a rating scale. A 5-point fully-labeled scale could start with "strongly disagree" or with "strongly agree." For rating scales with the same number of scale points and the same numerical and verbal labels, the key research question is whether varying the direction of the scale would affect survey responses, and if yes, in what way. Empirical evidence is mixed – some researchers showed empirically that the scale direction affects survey responses by pushing the answers to the start of a scale while others failed to witness a scale direction effect. This paper uses the 2008 American National Election Study Cross-Sectional data and the 2012 American National Election Study Cross-sectional data to examine the impact of question characteristics and rating scale design features on the scale direction effects. The goal of the paper is to identify question characteristics and scale features affecting the likelihood that a scale direction effect is observed on the resulted survey responses.

### **How Do You Measure Up?: Effects of Response Format on Life Satisfaction Measurement**

Francis M. Barlas, *GfK Custom Research*

Randall K. Thomas, *GfK Custom Research*

Most research on life quality has used bipolar scales (e.g., Satisfied to Dissatisfied), often showing a negative skew – people tend to respond with higher rather than lower responses, making comparisons more difficult due to ceiling effects. Both the polarity of the scale and extent of semantic labeling (fully anchored versus end anchored scales) have been shown to cause significant differences in results. The current study explored the use of scales across countries, comparing the bipolar versus unipolar scales and the fully-anchored versus end-anchored versions. We had opt-in respondents from different countries (Belgium, France, Germany, Italy, Spain, UK) participate in a web-based survey with questions on life satisfaction,

satisfaction with appearance, self-rated health, and comparative quality (if their lives were better or worse than 5 years ago). Respondents were randomly assigned to either a unipolar or bipolar seven category version of the response scale (the comparative scale was bipolar only – better/worse). Respondents were also randomly assigned to either have all scales fully anchored or end anchored.

For all three questions that had both a bipolar and unipolar variant (overall life satisfaction, appearance satisfaction, physical health), the fully-anchored unipolar scale was associated with lower mean values, but also seemed to show somewhat better discrimination across countries. In addition, both the unipolar and bipolar variants that were end-anchored scales showed little difference in how respondents used the scales. Results comparing countries can be significantly affected depending on the scale type, particularly with the fully anchored scales.

### **Modeling Anchoring Effects in Sequential Likert Scale Questions**

Marcin Hitczenko, *Federal Reserve Boston*

Surveys in many different research fields rely on sequences of Likert-scale questions to assess individuals' general attitudes toward a set of related topics. Most analyses of respondent results to such a series do not take into account the potential measurement error introduced by the context effect we dub “sequential anchoring.” Sequential anchoring occurs when the rating for one question influences the rating given to the following question by favoring similar ratings. Its presence can cause systematic bias in the study of relative ratings. We develop a latent-variable framework for question responses that capitalizes on different question orderings in the survey to identify the presence of sequential anchoring. We propose a parameter estimation algorithm and run simulations to test its effectiveness for different data-generating processes, sample sizes, and orderings. The model is applied to data in which eight payment instruments are rated on a five-point scale for each of six payment characteristic in the 2012 Survey of Consumer Payment Choices. We find consistent evidence of sequential anchoring, resulting in sizeable differences in properties of relative ratings for certain instruments. Finally, implications for future surveys and possible methodological solutions are discussed briefly.

## **Online Panel Research: A Data Quality Perspective**

### **The Status of Online Panel Research From a Data Quality Perspective**

Mario Callegaro, *Google UK*

Paul J. Lavrakas, *Independent Consultant*

Jon A. Krosnick, *Stanford University*

Online panels have become a prominent way to collect survey data. They are used in many fields including market research, social research, psychological research, election studies, and medical research.

Panel-based online survey research has grown steadily over the last decade. ESOMAR has estimated that global expenditures on online research as a percent of total expenditures on quantitative research grew from 19% in 2006 to 32% in 2012. In this panel session we will discuss data quality issues and the status of the art in research on online panels' data quality.

After many years of online panel research there is now cumulative evidence to look into the quality of data coming from online panels. Numerous associations and researchers conducted

comparison studies where the results from online panels are compared to benchmarks or other methods of data collection. We introduce a taxonomy on how to classify such studies and provide a brief summary of their outcome in the areas of point estimates, correlation among variables, and reproducibility of results. Finally we summarize the sparse literature on weighting online panels and its effect on data quality.

### **The Untold Story of Multi-Mode (Online and Mail) Consumer Panels: From Optimal Recruitment to Retention and Attrition**

Allan L. McCutcheon, *University of Nebraska-Lincoln*

In a recent study involving a mixed-mode experiment to recruit members to a consumer panel (Rao, Kaminska, and McCutcheon 2010), we tested the effect of various response-inducement techniques such as advance letters, monetary incentives, and phone follow-up on panel recruitment. The experiment was successful in recruiting individuals to a mixed-mode panel in which members do not receive any form of monetary incentives for their panel participation. Although the experiment used response inducements as a one-time stimulus to motivate individuals to join the panel, no such response inducements were provided for their subsequent panel participation. More than a year after the original experiment was conducted; we examine the question of how the propensity to continue participation in such mixed mode panels differ across the various treatments and demographic groups from which these individuals were recruited. In the current study, we draw upon the panel participation data to examine survival rates and the effect of individuals' demographic characteristics on the odds of staying in the panel. These findings have important implications for long-term panel participation.

### **Nonresponse and Attrition in a Probability-Based Online Panel for the General Population**

Peter Lugtig, *Utrecht University*

Since the mid-2000s, several research organizations have set up large probability-based online panel surveys. In these panels, individuals or households are followed over time to study change. When the respondents also represent the general population well, these panel studies can also be used to study change at the population level, or do cross-sectional surveys of the general population.

To make sure that these online panels can be used to generalize study findings to the population, they start out with a probability-based sample. After recruiting people offline, those without Internet access are then given access and a computer at home. In this way, the panel surveys will be representative of people with- and without Internet access.

Both initial nonresponse in the panel recruitment phase, and attrition over time threaten the external validity of probability-based online surveys however. This paper uses the Dutch LISS panel as an example to investigate the extent of nonresponse and attrition bias in the panel. To do so, we separate 9 groups of respondents, who each follow a distinct pattern of dropout. Then, within each group we look at the correlates of attrition, and compare these to the correlates of initial nonresponse bias. We show that initial nonresponse and attrition are two very different processes in a probability-based panel survey. The correlates of early attrition in the panel survey are very different from the correlates of initial nonresponse. We also find large differences between the correlates of different types of attrition, implying that attrition at various stages of the panel survey is selective.



In terms of the contribution to nonresponse bias, we find initial nonresponse bias to contribute more to overall nonresponse bias than attrition. The presentation concludes with a discussion of our findings and implications for survey practice.

### **Motives for Joining Nonprobability Online Panels and Their Association with Survey Participation Behavior**

Florian Keusch, *University of Michigan*

Because of the declining penetration rate of landline telephones and a general decline in the willingness to participate in surveys conducted through traditional data collection modes, survey data collection using online methods – especially online panels – is becoming more popular worldwide. Although these surveys have been widely adopted among marketing researchers, critics still fear that this sampling method leads to biased results produced by a breed of “expert” volunteer survey respondents who are solely interested in monetary incentives and therefore cannot be compared with the general population. This presentation first gives an overview of the existing literature on motives for participating in surveys in general and online panels in particular and how these motives influence survey participation behavior in the panel. Then, a new study among 1,612 members of an Austrian nonprobability online panel is presented. The results show that although other reasons for becoming an online panel member, such as helping to develop better products and services or entertainment, are reported more often, respondents who said that they joined the panel because of the promised monetary incentives have a higher starting and a lower break-off rate than do those who cited intrinsic reasons. Additionally, other characteristics of online panel members like the time of panel entry (panel tenure) and how they were recruited to the online panel seem to be at least as important as monetary motives in determining survey participation behavior in online panels.

### **Internet and Mobile Ratings Panels**

Paul J. Lavrakas, *Independent Consultant*

This study examines how Internet (PC and mobile) ratings panels are constructed, managed, and utilized. We provide an overview of the history and evolution of Internet/mobile ratings panels and examines the methodological challenges associated with creating and maintaining accurate and reliable Internet/mobile ratings panels. The research that has assessed the accuracy and validity of online panel data is critically discussed; as well as research that illustrates the type of scholarly and applied research questions that can be investigated using online ratings panel data. The presentation concludes with a discussion of the future of online ratings panels within the rapidly evolving field of Internet audience measurement.

## **Mode Effects**

### **I Know What You Did Last Sunday: An Investigation of Mode Differences in Reported Religious Behavior, Belief and Identity**

Daniel Cox, *Public Religion Research Institute*

Robert P. Jones, *Public Religion Research Institute*

Juhem Navarro-Rivera, *Public Religion Research Institute*

As more public opinion surveys are moving from telephone to online modes, it is imperative that researchers operate with a complete understanding of how mode choices affect responses,

particularly on questions that are susceptible to social desirability effect. A growing body of research has documented that, on many important policy questions and political attitudes, there are few systematic differences between modes. Based on these initial results, many researchers, including prominent organizations such as the Associated Press, have switched from RDD telephone surveys to probability-based online panel surveys. However, more work is needed to fully what measures are potentially impacted in the shift from live-interviewer telephone surveys to self-administered online surveys.

Recent research conducted by Public Religion Research Institute affords us the unique opportunity to investigate the question of social desirability bias in reporting not only of religious behavior but religious beliefs and identity as well. In late 2013, PRRI conducted two separate surveys—one using traditional telephone methods with live interviewers, while the other utilized an online survey design based on GfK's Knowledge Panel. Each survey was based on a random sample of adults (age 18 and up) and included identically worded questions about the frequency of worship attendance, belief in God, salience of religion, and religious identity.

This paper will examine to what extent Americans respond to basic religious questions differently across different modes. Our expectation is that Americans will over-report their frequency of religious attendance in telephone surveys because of latent pressure to provide a socially desirable answer. We expect there to be less systematic inflation of religious beliefs and identity. Finally, we will explore the implications of our findings for our understanding America's religious profile and the ways in which basic religious measures like religious attendance are used in analyzing more general polling on issues and politics.

## **Computing Response Rates for Mixed-Mode Survey Designs**

Ana Villar, *City University London*

Mario Callegaro, *Google UK*

There is good documentation and formulas on how to compute response rates for single-mode surveys, specifically, face-to-face, telephone, mail and web surveys. AAPOR has produced numerous versions of the "Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys" that addresses the issue of computing response rates in a standardized way. However, despite the increasing use of mixed-mode designs in survey research, computation of response rates for such designs is only briefly mentioned in the AAPOR standards. In this paper we propose a way to compute response rates when mixed-mode surveys are used. Starting from the AAPOR definitions, we provide a discussion on the implications of recording contact dispositions and computing response rates for different types of mixed-mode designs. We argue that the key factor affecting the computation of response rates in mixed-mode surveys is whether more than one mode is used to contact respondents (rather than to collect data), and that certain mixed-mode designs can in practice apply single-mode outcome rate computation procedures. We then present a proposal for computing response rates in mixed modes and discuss the implications for sequential and concurrent mixed-mode designs, considering mode-specific rates for each contact mode as well as combined rates for entire survey. However, given that nonresponse components and eligibility rates vary across modes, computing e for a combined response rate poses particular challenges. We present and compare two different ways to handle this. As an example of the complexities associated with documenting and reporting response rates in mixed mode designs, we use the disposition codes of the European Social Survey (ESS) mail/web mixed mode experiments conducted in 2012 in Estonia, Sweden and the UK, and we compute response rates for these mixed mode surveys.

## **Impact of Mode Design on Measurement Errors and Estimates of Individual Change**

Alexandru Cernat, *University of Essex, Institute for Social and Economic Research*

Mixed modes are receiving increased interest from survey methodologists as a possible solution to saving costs while retaining high quality data. In recent years this interest has extended also to panel studies which are looking to save costs by including a cheaper mode for some of their respondents.

The current presentation aims to tackle some of the issues linked to such a design. First, I aim to see if using a mixed mode design will increase systematic and random error compared to a single mode CAPI survey by applying equivalence testing in a Confirmatory Factor Analysis. Secondly, I aim to investigate if estimates of individual change are influenced by mode design by comparing latent growth models across the two groups.

The first four waves of the Innovation Panel, part of Understanding Society (UKHLS), will be used for the analysis. The second wave of the data randomized respondents to either a single-mode CAPI design or to a CATI-CAPI sequential design. The SF12 health scale will be used to investigate both measurement equivalence and estimates of individual change while individual and household characteristics from wave one will be used for the propensity score matching.

**Sunday, May 18**  
**9:45 a.m. – 11:15 a.m.**  
**Concurrent Session K**

## **Interviewer Effects on Data Quality**

### **Changes in Interviewer-Related Error Over the Course of the Field Period: An Empirical Examination Using Paradata**

Antje Kirchner, *University of Nebraska-Lincoln*

Kristen Olson, *University of Nebraska-Lincoln*

It is well-established that interviewers learn behaviors both during training and on the job. However, how this learning occurs has received little empirical attention. There are two competing hypotheses about what happens during field data collection—one is that interviewers learn behaviors from their previous interviews, and thus change the measurement situation in reaction to the behaviors previously encountered. We refer to this as the contagion hypothesis.

The second hypothesis is that interviewers encounter less cooperative respondents as the data collection period progresses, and thus change the measurement situation in reaction to these types of respondents (i.e., nonresponse propensity affecting the measurement error situation). We refer to this second hypothesis as the composition hypothesis.

This paper examines these two hypotheses using data and paradata from the Work and Leisure Today survey (AAPOR RR1=4.2%). We use multilevel models to account for the nesting of respondents within interviewers. There is a significant interviewer variance component on both the length of the interview (ICC=0.22) and the number of answer changes in the survey interview (ICC=0.09).

Initial results support the contagion hypothesis over the composition hypothesis for the length of interview. Consistent with previous research, experienced interviewers generally have significantly shorter interviews than inexperienced interviewers. As interviewers gain study-specific experience, the length of interview decreases, but this effect is less pronounced for experienced interviewers. We find no difference in length of interview for respondents with higher or lower response propensity. Our results are less conclusive for the number of answer changes and do not lend support to either hypothesis. The paper ends with a discussion of the implications of the findings for interviewer training and monitoring.

### **What Factors Explain Variation in Monitors' Detection of Interviewing Errors in Telephone Surveys?**

Douglas B. Currivan, *RTI International*

Paul P. Biemer, *RTI International*

Tamara Terry, *RTI International*

Ivan Carrillo-Garcia, *RTI International*

To limit the potential for interviewer behavior to bias or add variance to survey estimates, monitoring telephone surveys requires accurate and consistent detection of interviewing errors.

Multiple factors can affect monitors' detection of interviewing errors, including monitors' characteristics and attitudes, how interviewers are supervised, respondent characteristics and behavior, interviewer characteristics, the survey questions and protocol being monitored, the telephone survey center environment, and the monitoring system being used. RTI has developed a standardized, mode-independent interview quality monitoring evaluation system (QUEST) that allows telephone interviewing behaviors to be evaluated using a common set of quality metrics based on empirically derived indicators. The standardized, empirical approach used to design QUEST was intended to promote accuracy and consistency in telephone monitors' detection of interviewing errors.

Because QUEST data alone are insufficient to understand how multiple factors contribute to variability in telephone monitors' error detection, this research will combine data from (1) QUEST monitoring sessions, (2) information provided by monitors at the beginning of sessions, (3) responses to a survey of monitors, and (4) administrative records with interviewer characteristics. These data will allow for operationalization of multiple sources of variation in monitors' error detection that cannot be captured in the monitoring system and results. For example, the information provided by monitors in QUEST and their survey responses will provide data on how situational factors such as interview complexity or challenging respondents might affect the number, type, and severity of errors monitors record in a session. Multi-level models will be used to analyze the contribution of these various sources to variation in monitors' detection of interviewing errors. This presentation will discuss the implications of these results for understanding sources of variation in monitors' detection of interviewing errors and for guiding decisions on designing monitoring processes in centralized telephone survey centers.

## **Questionnaire Design in Telephone Surveys: Interviewers' and Call Center Managers' Experience**

Wojciech Jablonski, *University of Lodz*

Telephone interviewers are much more constrained than face-to-face interviewers as far as using visual aids is concerned. Lack of visual materials is usually compensated by the implementation of special techniques while designing the interview script. Such techniques include, for instance, using semantic scales instead of numeric ones, unfolding questions based on semantic scales, using fewer answer categories and shortening the answer categories, as well as formulating pre-categorised questions. Moreover, as communication between the CATI interviewer and the respondent lacks the non-verbal component, telephone interviews need to be shorter than face-to-face interviews.

Within the presentation, we will outline the results of the study what was carried out in 2009 and 2010 among 12 major Polish commercial survey organizations. The research was based on standardized self-administered questionnaires for CATI interviewers, as well as in-depth interviews conducted with well-experienced interviewers. A total of 846 questionnaires and 32 IDIs were completed. Additionally, in 2013 a follow-up research was carried out. During IDIs CATI survey managers from different companies were asked to comment on the results obtained in both quantitative and qualitative studies.

This presentation investigates the issue of interviewers' and call center managers' opinions and attitudes connected with different CATI questionnaire designs (response scales formats in particular) used by survey organizations. The interviewers were encouraged to elaborate on methodological solutions implemented in CATI scripts, as well as to describe problems they encounter while asking different types of questions. Research call center managers in turn

commented on the interviewers' observations, pointing out the factors that may lead to fieldwork difficulties.

As we see it, CATI call center staff opinions are valuable sources of information about the interview process, including the design aspects of the questionnaire and their impact on data reliability. These perspectives should be taken into consideration while preparing research tools used in telephone surveys.

### **Exploring the Implications of Interviewer Behavior on Data Quality**

Rodney L. Terry, *U.S. Census Bureau*

Ryan King, *U.S. Census Bureau*

Laurie Schwede, *U.S. Census Bureau*

A tenet of interview standardization is that in order to collect data that is as reliable as possible, interviewers should read questions exactly as worded. To that end, behavior coding is a questionnaire evaluation method used to analyze interviewer and respondent interactions to examine the extent to which interviewers properly read questions as worded, as well as the extent to which respondents give appropriate (i.e., codable) answers. Previous U.S. Census Bureau research has found that some interviewers do not read questions as worded, and at times do so at high percentages. However, these studies do not directly assess the relationship between interview behavior coding data and interview data quality. The present study's purpose is to explore this relationship.

We use data from a Census Bureau comparative ethnographic evaluation of the 2010 Census Coverage Measurement (CCM) operation and analyze portions of the interview in which interviewers asked respondents demographic questions. The 2010 Census CCM evaluation was conducted at nine targeted race/ethnic group sites in the United States. At each site, an ethnographer with expertise on the targeted race/ethnic group accompanied Census Bureau interviewers to systematically observe 318 live, face-to-face interviews. With respondent permission, these interviews were tape-recorded. Later, about half of the interviews were behavior coded. For 158 of the interviews, we analyze whether interviewer behavior codes for non-ideal behavior are associated with indicators of poor data quality, including missing data, response refusals, and repeated-measurement inconsistency between matched CCM interview and 2010 Census data. We then discuss how the results may help focus questionnaire design and testing resources to questions that – when read in a non-ideal way – have the greatest impact on data quality.

### **The Effect of Interviewer Probing on Item Nonresponse and Measurement Error in Cross-National Surveys: A Multi-Group Latent Variable Analysis**

Sarah Butt, *City University London*

Jouni Kuha, *London School of Economics*

Chris Skinner, *London School of Economics*

Myrsini Katsikatsou, *London School of Economics*

Item nonresponse in attitudinal surveys can lead to a loss of efficiency and introduce additional sources of measurement error affecting the validity of conclusions drawn from the data. One possible solution is for interviewers to probe “don’t know” (DK) responses, encouraging respondents to give a substantive answer if possible. Probing may elicit meaningful responses from initial DK respondents and thus improve overall data quality. However, probing may also introduce new measurement error because it pressures respondents into returning poorly

considered answers. This is a particular concern for cross-national surveys such as the European Social Survey (ESS) as different responses to probing between countries may distort the validity of using the data for country comparisons.

This paper presents the results of an experiment designed to test the effect of interviewer probing of DK responses on data quality across different European countries. As part of the ESS Innovation Sample, in late 2013 respondents in Hungary, Bulgaria and Portugal were asked questions on attitudes to welfare provision via face-to-face omnibus surveys. Three-quarters of respondents were prompted to give another response in the event that they said DK initially whilst the rest, acting as a control group, were not probed. Multi-group latent variable models specifically designed to take account of measurement differences between groups will be used to determine the effect of probing on levels of nonresponse, and whether responses generated by probing differ from unprobed responses. The technique will also be used to compare the effect of probing across countries, different types of respondent and different types of questions (standard agree/disagree items vs. items designed to evaluate knowledge). Findings from the experiment will be used to make recommendations regarding the use of probing on future surveys.

## **The Impact of Mass Media on Public Opinion**

### **All News is Not Equal: Analyzing the Impact of News Sources on Presidential Approval**

Dan Cassino, *Fairleigh Dickinson University's PublicMind Poll*

Using a comprehensive dataset of human content-analyzed news coverage from television news broadcasts on ABC, NBC, CBS and Fox News, the study analyzes the differential impact of messages about Barack Obama on approval ratings among various demographic groups weekly from 2009 to 2012. The results demonstrate that media messages about Obama drove changes in public opinion, rather than the other way around, and that various groups respond very differently to these messages, especially based on party and ideology. The results are also used to demonstrate the outsize effect of Fox News on American public opinion, especially among voters who identify as conservative: while media tenor towards Obama, in general, is a significant leading predictor of aggregate public opinion change, only Fox News has an independent effect on aggregate public opinion. Even more interesting, the difference in tenor between Fox and the other news sources is itself a leading predictor of approval change among some groups, providing some evidence for a motivated reasoning-based approach to these shifts in approval. The data is also used to create a short-term forecasting model for presidential approval based on current media coverage.

### **Gun Control and the Press After the Sandy Hook Killings: The Relationship Between Leading National Newspaper Coverage, Public Opinion and Public Policy**

Thomas B. Christie, *University of Texas-Arlington*

This study examines leading national media coverage of prominent issues in the public policy debate over gun control in the aftermath of the Sandy Hook Elementary School tragedy in Newtown, Conn., Dec. 14, 2012. This media coverage is viewed in light of the shifting public support of national legislation culminating with the rejection of several measures related to gun control by April 17, 2013.

As the nature and extent of media coverage often precedes the formation of public opinion and public policy in a democracy, this study uses the agenda-setting and agenda-building traditions of mass communication research as a basis for understanding the linkage of gun control issues between news media, public policy and public opinion. It examines the nature and extent of this coverage in light of the national focus on the issue, followed by waning support of gun control legislation as found in public opinion polling (CBS News, 2013) and majority public agreement on some gun control issues (Pew, 2013). Key research questions driving this study were:

1. How were the issues of the gun control/violence defined by three major national newspapers leading up to the 2013 U.S. Senate votes?
2. How did news media coverage of gun control/violence issues contrast with issues noted in public opinion polling during the same time period?
3. What are the differences in coverage of the gun control/violence issue among the three major national newspapers examined?

The study begins with the news media coverage of the issue in the immediate aftermath of the tragedy on Dec. 14, 2012, and concludes after legislation was defeated in the U.S. Senate on April 17, 2013. Results illuminate how the issue was covered and reveal major differences in the nature of coverage among The Wall Street Journal, The New York Times, and Washington Post.

### **“Powerful and Emotional” Pictures in Television News: Effects on Viewer Comprehension and Political Evaluation**

Erik P. Bucy, *Texas Tech University*

Jacob Groshek, *Boston University*

During the 1984 presidential election Leslie Stahl aired an infamous, 5-minute report on the CBS Evening News that has come to define “audiovisual redundancy”—the degree of congruity between the audio and visual tracks of broadcast news. In her report, Stahl was critical of the Reagan administration’s campaign stagecraft in her spoken narration. Crucially, however, this critical commentary was framed against a highly positive visual backdrop: shots of the candidate smiling and shaking hands with supporters, admiring crowds, patriotic symbolism, and other positive imagery that appeared pro-Reagan. We examine the veracity of Dick Darman’s comments that, “when the pictures are powerful and emotional, they override if not completely drown out the sound” (Stahl, 1999). A between-subjects experiment involving 242 undergraduate participants was designed to test this proposition, known as the visual primacy effect. Specifically, we assess how individuals respond to different forms of audio and visual information by controlling the conditions by which political news is presented (written transcript, audio only, video only, and full audio/video). Data analysis finds conditional support for Darman’s hunch in terms of story comprehension, information recall, affective response, and political evaluations.

In addition, we ran a series of six focus groups to qualitatively assess viewer responses to such “memorable moments in televised politics.” Focus group comments indicate: recognition of a gap between political image and reality but respect for Reagan as a political performer, or shaper of political reality; awareness of visual framing but also political image management that makes framing necessary; visual/verbal counterbalancing—holding positive feeling towards Reagan (who exudes a “feel good factor”) even while recognizing and agreeing with the criticism in Stahl’s report; and, ambivalent feelings towards the press for seeming to be



biased/critical but showing Reagan in a positive light (leading to a sense of reinforcement for supporters and critics alike).

## **Diverging State and Nation: Correspondence Between Statewide and National Election Surveys in 2012**

Josh Pasek, *University of Michigan*

Increasingly, scholars and polling aggregation websites are combining state and national level survey data to better estimate election outcomes. The models used by sites such as Pollster and FiveThirtyEight typically presume that the processes generating state and national data vary in similar ways. The current study tests the accuracy of this assumption. Gathering data from 1775 sample surveys conducted during the 2012 election, this study assesses how closely trends in survey estimates from individual states anticipate corresponding trends in national estimates as well as survey estimates in other states. In general, we find that estimates from individual states and national level estimates do trend together, but not as strongly as might be expected. Over-time variations in some states, such as Pennsylvania, were not even associated with national trends. The correspondence between states only became a little stronger when accounting for mode and house effects. Hence, the results seem to imply either that states differ notably from one-another or that campaign effects could be introducing important variations into state-by-state election tracking that are not fully accounted for in national-level surveys.

## **The Newspaper Shook My Confidence and the County Clerk Broke My Trust: Media Effects and Interactions with Local Authorities in Building and Deteriorating Public Confidence in Democratic Institutions**

Dzmitry Yuran, *University of Tennessee*

Melissa W. Graham, *Oklahoma City University*

Michael R. Fitzgerald, *University of Tennessee-Knoxville*

Public trust and confidence are necessary lubricants in the mechanisms of most modern societies. Democratic institutions and systems in which they are embedded, function successfully only when the publics they serve agree to rely on them. Understanding the sources of these lubricants, as well as factors eroding them, is crucial in the discussion of the development and operation of democracy as a phenomenon. Mass media arguably play a very significant role in creating images of the political world and its elements in the public mind. Walter Lippmann's "out of sight out of mind" mantra is being repeated over and over again by media effects scholars – if we don't deal with things directly, we have to rely on media to draw a picture for us. A significant body of research illustrates a direct (and negative) connection between amounts of media exposure and levels of trust and confidence among audiences. Trust in media themselves appears to correlate with confidence in democratic institutions as well. Other factors may be just as significant in the confidence equation. There are reasons to believe that democratic systems at large, as well as the individual institutions composing them, may often be 'guilty by association.' That is, all institutions are diminished in the public mind due to negative first-hand experience with any given institution, which in most political systems are likely to occur at the local level of government—most especially interactions with local authorities. This study examines the relationship between mass media exposure and contact with local authorities as factors associated with public confidence in democratic systems through the analysis of survey and experimental data. In order to do so, it develops scales for overall levels of confidence in democratic institutions among publics as well as the quality of their relations with local authorities.

## Research on Non-Probability Samples

### A Model-Based Approach for Achieving a Representative Sample

George Terhanian, *Toluna*

John Bremer, *Toluna*

Carol Haney, *Toluna*

AAPOR issued a task force report recently that focused on non-probability sampling. The report acknowledges that it is perhaps more difficult today than ever before to conduct any mode of survey research when employing population-wide probability sampling.

The AAPOR report is far more descriptive than prescriptive, offering several conclusions but no explicit recommendations on next steps. We believe this may be an oversight, not least because immediate opportunities are available for survey researchers to learn more about non-probability sampling. For instance, the ARF mounted a new large-scale research project in January, 2013 to assess, among other objectives, the comparative effectiveness of (a) several methods—probability and non-probability—of selecting individuals for surveys, and (b) several different approaches of making post hoc adjustments to those data.

The aim here is to rely on data from the ARF study to develop an optimal model—by that we mean one which minimizes sample bias, thereby resulting in a representative sample. Our plan is to proceed as follows:

1. Identify benchmarks from the ARF study, and others sources, that represent, in some sense, the “truth,” or the standard against which we will measure bias.
2. Develop simple and more complex statistical models for selecting respondents through “sample matching,” which the AAPOR report identifies as “the most promising non-probability approach for surveys.”
3. Utilize standard optimization techniques to examine whether and to what extent different methods of post hoc adjustment, including propensity score adjustment, reduce bias.
4. Test the developed respondent selection and weighting models with new data, recognizing that there may be competing explanations for why these models are more or less effective than suggested by the original simulation.

Through this process, we believe more will be learned about how and why non-probability sampling can be a dependable, credible source of information and evidence.

### Comparing the Results of a Survey of Physicians Using Probability and Non-Probability Samples

Emily Geisen, *RTI International*

Murrey Olmsted, *RTI International*

Lily Peng, *RTI International*

Physicians are an elite population and are often surveyed in public health research. As a result, achieving high response rates in probability-based sample physician surveys can be difficult and costly. In an effort to explore a lower cost alternative we examine whether a non-probability sample of physicians yields different results compared to two probability samples of physicians. One of the probability samples was selected from the American Medical Association (AMA) Masterfile. The second probability sample was selected from the Dximity Physician Masterfile. Both sources compile data on physicians from medical school graduate records and US medical

licensure data. The non-probability sample was selected from members of Doximity's professional network – a fast growing communications and social media network for doctors that includes over 200,000 physician members.

We first compared the demographic distributions of the three sample frames. We found that with respect to age, gender, and region the three samples were generally comparable. We then administered an identical survey to physicians from each of the three samples. A web and mail mixed mode approach was used for the probability samples and a web-only approach was used for the non-probability sample.

In this paper we will determine whether there are differences in the type of physicians who responded to the three surveys and whether these samples differed on the survey estimates. For any differences, we will explore whether these are due to mode differences, response bias, or other unobserved factors. Finally, we will explore whether data quality could be improved through the use of weighting of the non-probability survey to national benchmarks.

### **Are Citizen Surveys Conducted via Opt-In Web Ready for Prime Time?**

Thomas Miller, *National Research Center, Inc.*

Erin Caldwell, *National Research Center, Inc.*

From a probability sample using a USPS sampling frame and multi-contact collection method in more than 40 jurisdictions spread across the U.S., a random sample of residents completed The National Citizen Survey which asks standard questions about quality of community life, service delivery and citizen engagement. Unique policy questions are also part of most surveys. As soon as data collection by mail had closed, the same survey was posted on community websites with an offer for any resident to respond. This paper examines the differences in the sample sizes, demographics, responses and recommendations for policy that derived from respondents of the probability sample by mail compared to the opt-in web survey takers -with and without post-hoc weighting - and it makes recommendations about the viability of opt-in Web surveys for uses in local government planning, budgeting, performance measurement and policy evaluation.

### **Towards a New Math for Non-Probability Sampling Alternatives**

Mansour Fahimi, *GfK*

Wendy Gross, *GfK*

Frances Barlas, *GfK*

Traditional methods of survey research that rely on Neyman's probability-based sampling paradigm are facing increasing challenges from several perspectives. On the one hand, common methods such as address-based sampling or dual-frame random digit dialing are subject to coverage issues that may not be fully ameliorated through cost-effective remedies. On the other, response rates continue to deteriorate for all surveys, even when resource-intensive refusal conversion strategies are employed. Add in the growing need for cost containments, and it is no wonder why alternative sampling methods are gaining popularity, even among traditional survey research scientists. Top among such alternatives are samples that are secured from online panels, commonly referred to as opt-in samples. Due to lack of selection probabilities, however, measurable statistical inferences cannot be generated from such samples. The authors will review a number of practices that are currently used for developing robust inferences from opt-in samples, as well as discuss refinements that could further reduce some of the biases that are inherent to such samples.

## **Assessing the Performance of Non-Probability Online Surveys on Non-Electoral Behavioral Measures**

Clifford A. Young, *Ipsos Public Affairs*

The survey research world is changing. Gold standard methodologies such as the telephone survey are under increasing pressures due to declining response rates, increased cell phone-only households, and rising costs. Many have argued that one possible or partial solution to this problem is the online survey. There is some evidence of this. Indeed, as a class, online polls performed well at the 2012 US presidential elections. However, online polls have their serious detractors as well. One criticism is that online surveys might perform well in the “highly modeled” context of elections, but less so when it comes to other behavioral and nonpolitical measures. With this in mind, we ask one simple research question in this paper: how do online non-probability samples perform on non-electoral behavioral measures?

To do this, we will analyze approximately 280,000 interviews collected for the Reuters/Ipsos tracking poll between November 2011 and March 2014 — twenty-nine (29) months in total. The Reuters/Ipsos poll is a blended online sample where multiple panel and non-panel sample sources are combined.

Our paper will focus on the performance of non-probability polls across a dozen different behavioral measures, all with official external benchmarks from HHS, DOD, CPS, BLS, or the Census. Some benchmark measures include: percent active and reserve military, percent of uninsured Americans, percent uptake of the winter Flu Shot, proportion of foreclosed Americans, among others.

To measure performance, we will employ a ‘Mean Square Error’ framework looking both at bias (average absolute difference) as well as variability of the estimate. Each month will be treated as a separate data point. This together with 12 different benchmark measures gives us 348 separate observations for our analysis. Finally, as a secondary objective, our paper will examine the performance of difference calibration models, including propensity matching as well as demographic and attitudinal post-stratification weighting.

## **New Developments in Data Collection with Smart Phones and Mobile Devices**

### **You CAN Touch This: An Experiment to Compare Computer and Mobile Surveys Using Touch Friendly Question Types**

Joey Michaud, *Maritz Marketing Research*

Trent D. Buskirk, *Marketing Systems Group*

Ted Saunders, *Maritz Marketing Research*

Smartphone ownership continues to rise in the U.S. and email and internet activity on these devices is beginning to surpass such activity on either desktop or laptop computers. The consequence for online surveys is the continued rise in unintentional mobile completes, with some sources citing that at least one in five web surveys are being accessed via mobile devices. While the basic research into optimizing the survey experience and data collection on mobile devices is growing there are still fundamental gaps in our knowledge of how to optimize

certain types of questions in the mobile setting. Specifically, survey researchers are still trying to understand which online design principles directly translate into presentation on mobile devices and which principles have to be modified to incorporate separate methods for these devices. In this research we will explore various question types that take advantage of the mobile (and touch) platform and compare how data collected on mobile devices compares to that gathered on a computer using similar question types. The question types we will explore are free response questions with prompting for voice to text input, slider scales with various start positions including off-scale, and the use of roller bars compared to standard drop-downs for different numbers of response options. Panelists will be randomly assigned to one of several possible experimental conditions using a factorial design within each of three device type blocks – computer users, tablet users and smartphone users. Factors to be manipulated include standard/slider scales, slider position, number of scale points, scale numbering, and the number of list items. We will compare key metrics including: response distributions, completion rates and times, item response rates, character counts, and the use of horizontal repositioning. Respondents will also be asked directly about their preferences for response options and their satisfaction with the survey.

### **Nonresponse in a Mobile-Web Survey: a First Look at the Causes and the Performance of Different Predictive Models**

Christopher Antoun, *University of Michigan, Institute for Social Research*

A growing number of people are responding to Web surveys on their phones. This emerging method of survey data collection, which is commonly referred to as “mobile Web,” provides new opportunities for survey researchers. For example, by enabling respondents to complete surveys where it is convenient, mobile-Web surveys could have higher participation than other modes. Yet, early research suggests that people are less willing to participate in mobile-Web surveys than traditional Web surveys. Why is this? The current study explores the mechanisms, or causes, of nonresponse in a mobile-Web survey. 1263 members of the Longitudinal Internet Studies for the Social Sciences (LISS) panel completed a traditional Web survey about attitudes toward and use of technology. Next, these same panel members were invited to participate in a mobile-Web survey, and we observed who responded and who did not. Nonresponse propensity models were developed using both demographic characteristics from the LISS panel frame and technology measures from our baseline survey. We first explore the correlates of nonresponse. Even after controlling for demographic differences in response rates, we expect to find that technology measures, such as type of phone and frequency of mobile-Internet usage, will predict participation. Then we compare the performance of two propensity modeling techniques: we investigate whether a data mining program called SEARCH improves prediction over logistic regression, a more conventional method. This comparison will shed light on whether complex interactions among technology variables that can more easily be found using a data mining procedure should be considered in response propensity models for mobile-Web surveys. These findings will help add to an emerging picture of why people do or do not participate in Web surveys using their phones.

### **Nonresponse and Measurement Differences in Mobile vs. Traditional Online Surveying: Findings from Two Experiments**

Paul J. Lavrakas, *Independent Consultant*

Robert Clancy, *Usamp*

Two large national experiments were conducted in 2013 gathering data from uSamp’s general population panel members that all have agreed to respond to both traditional online surveys and

mobile device surveys. For each survey a random sample of 30,000 panelists was drawn and randomly divided into two groups. The two groups were compared after the randomization to assure their demographic equivalence. One group was invited to participate in an online survey; the other group was invited to participate in a mobile survey. For each experiment, equivalent questionnaires were used for the online and mobile conditions. In the first study the topic of the survey was shopping behaviors; in the second, the topic was current events. The first study was completed in the spring of 2013; the second in the fall of 2013. Findings from the first study indicated significant differences in the amount of nonresponse between the two conditions, with the online survey condition having the higher response rate, but no appreciable differences between the two survey modes was found in nonresponse bias. However, data quality was significantly higher in the mobile survey condition, with less recall error and social desirability error than in the online condition. The data from the second survey have not been analyzed. Discussion will include attention to the issue of when the mobile survey mode is more cost-effective to use compared to a traditional online mode.

### **Scan All: Smartphones for Measuring Household Purchases in Developing Markets**

Jeff Scagnelli, *The Nielsen Company*

Kelly Bristol, *The Nielsen Company*

Developing countries are of growing importance within the Consumer Goods industry, and there is a need to provide better insights into the consumption of household goods through more accurate measurement. Leveraging the mobile infrastructure in these areas allows household purchases to be captured in a more immediate fashion which reduces recall bias. Mobile devices also offer the ability to combine user responses with passive measures of behavior for a complete picture of the user experience. With the growth in this area there is a need to better understand the participation, and tolerance for using mobile devices over time to gather consumer behavior longitudinally.

This paper contributes to the gap in literature regarding mobile barcode scanning of consumer products. We will examine the results of a pilot study conducted in South Africa where users report the consumption of household goods through an Android smartphone app. Households were provided with a smartphone for reporting purchases made, which they could also use for phone calls, texting, and data. Households were instructed to report all consumption of food, beverage and household staples via the app. With barcode scanning being the primary mode of collection, we will review in greater detail the outcome of panelist scanning activities, as well as the frequency of reporting during the study. These insights will add to the growing knowledge base of mobile data with an international context.

### **To App or Not to App: The Key Questions to Ask Before Considering Mobile & Web Application Measurement**

Jennie W. Lai, *Google*

Michael W. Link, *The Nielsen Company*

Kelly Bristol, *The Nielsen Company*

Shu Duan, *The Nielsen Company*

With over 80% of adults in the U.S. using the Internet and half of all adults being smartphone owners, there is increasing interest on mobile & Web app measurement as the emerging mode of data collection. Before the next wave of methodological research on important questions such as coverage, data quality, mode effect, nonresponse, etc., there should be careful evaluation

whether application measurement is the right method for research purposes. App development is not an inconsequential undertaking considering the significant cost, time and infrastructure resources required. This approach is arguably best suited for ongoing measurement services to supplement an existing mode of data collection for longer-term data collection period. Keeping in mind the measurement objective, this research reports two key areas of considerations for 1) mobile and/or Web app platform for data collection such as usage statistics across operating systems; and 2) usability design of the app for sustaining engagement such as notification methods, instructional tutorials, monetary/non-monetary incentives, etc. These insights are gathered from Nielsen's latest 3.0 version of the mobile & Web app to collect television viewing behavior through an electronic diary method. After the two-week data collection period, these respondents were also asked to complete a Web survey to follow up on their study experience. The dawn of mobile & Web app measurement is upon us with new possibilities to explore and measurement questions to address. This research paper will address the feasibility of an app as a measurement tool and user profile by types and features of the tool.

## **Geocoded Data Applications**

### **Does Appending Billing Zip Code Make Cents?**

Missy Mosher, SS/

Edward P. Johnson, SS/

Linda B. Piekarski, SS/

Starting last year the major suppliers of telephone sample added the ability to target down to the zip code level on the Wireless RDD sample frame by appending the billing zip code. Although wireless phone numbers can't be reliably linked to a residential address like numbers from the landline frame, the additional append of billing zip code to the sample has proven effective in small geographic targeting. However, on average only 40-45% of a given sample will be linked to a billing zip code and the percentage can vary greatly by geography. Last year, Abt SRBI conducted a study where they found significant differences in a number of key characteristics between the zip and non-zip sample strata. The sample that did not have a billing zip code appended tended to be less educated, lower income, more minorities, and renters. They also found differences in risky driving behavior (the topic of their study) and other key behavioral metrics. Clearly, the sample with no zip code cannot be ignored.

We extend this research to examine a different topic (in this case health, auto, and homeowners/renters insurance) and a different geography (a group of zip codes in Des Moines, IA). We also look at the interviewer cost savings associated with collecting 75% of the sample from records with the billing zip code and 25% of the sample from records without a billing zip code. We will compare the increase in the margin of error due to weighting to the decrease in the margin of error from a larger sample size because of the better targeting to see if the targeting was worth the money. Lastly, we will report on representation level in the insurance market as well as general attitudes towards insurance and risk management from the respondents who had a zip appended.

## **Understanding Mobility: Consent and Capture of Geolocation Data in Web Surveys**

Scott D. Crawford, *Survey Sciences Group, LLC*

Colleen McClain, *University of Michigan*

Toben Nelson, *University of Minnesota*

Robert Young III, *Survey Sciences Group, LLC*

The use of smartphones and tablets of various sizes and operating systems has posed numerous challenges for researchers, impacting survey response and data quality. Importantly, though, their use also provides opportunities to capture nuanced data helpful in understanding their impact. A key element of such device use is the mobility of the respondent; one may be in transit or in a public location while responding to a Web survey previously only completed on a computer. Thus far, little data has been collected to assess the impact of mobility on data quality. With these aims in mind, we report the findings of an experiment assessing feasibility of collecting geolocation from survey respondents, as well as the quality of data received. In doing so, we discuss implications for survey implementation and measurement error in an increasingly mobile world.

Within a series of short web surveys of college students fielded in 2013, we assessed feasibility by asking respondents whether they would be willing to let researchers collect geolocation data in the future. Building off positive results from this analysis, we then experimentally tested methods of informed consent for geolocation capture. Respondents were randomly assigned to receive no request for geolocation, a request via the survey consent form, or a request via a separate question focusing only on this data capture. We captured geolocation data at every page of the survey for those consenting.

We will discuss the results of the experiment, laying out a recommendation for obtaining consent; comparing those who consented to those who did not; and assessing the quality of the geolocation data obtained. We will also present analyses examining how such information may be used in a broader context-- exploring whether or not the mobility of the respondent may be a key predictor in evaluating survey data quality.

## **All Survey Data Are Spatial: Practical Tips on Geocoding and Performing Distance Calculations**

Timothy B. Gravelle, *University of Essex & PriceMetrix Inc.*

Numerous research questions in applied social science and market research bear on the location of survey respondents, and by extension their proximity or distance to geographic features of interest (e.g., international borders, oil pipelines, retail locations, etc.). Research can be designed in the first instance to explore such spatial dynamics, but sometimes with secondary data analyses, one has to "do the best with what you've got," relying on whatever geographic indicators happen to have been retained in a dataset and the statistical software one already has (meaning no access to specialized geographic information systems (GIS) software). This methodological brief has such a practical orientation, and describes how various data points (ZIP codes, telephone numbers, street addresses, census geographic codes) can be used to append latitude-longitude coordinates to survey data (geocoding) using a range of commercial and freely-available resources. It also discusses different methods of calculating distances using standard statistical software.



## **Multi-Level Models: Connecting Geographic Information to Individual Level Survey Responses**

Kenneth Fernandez, *Elon University*

Jason Husser, *Elon University*

Robert Huckfeldt (1986, 1) argued that “political opinions and behavior of individuals cannot be understood separately from the environment within which they occur.” Over twenty-years later incorporating multi-level models is not uncommon, but is still far from standard practice in many fields. Including variables that capture geographic characteristics (e.g., percent minority, homicide rates, unemployment rates) can greatly improve our understanding of public opinion and increase our predictive capability, but there are still challenges to the analysis of such multi-level models. Currently, there is no consensus as to the most appropriate statistical techniques to use in examining a model that includes individual-level variables as well as geographic/contextual variables. The paper reviews the benefits and challenges of multi-level models, conducts a systematic review of how contextual variables are treated in the social sciences, and then conducts a replication study of three statistical models. The research project shows how attitudes toward immigration are greatly influenced by contextual factors including the size and growth of the Latino population, the partisan balance of a community, and the economic characteristics of an area (unemployment rate, agricultural versus manufacturing community). The paper then compares statistical findings from three models: a hierarchical linear model, a fixed-effects model using dummy variables, and a traditional regression model using robust standard errors correcting for clustering of respondents.

## **Efficient Sampling Design and Screening in Surveying Rare Population in Small Areas**

Chintan Turakhia, *Abt SRBI, Inc.*

Marci Schalk, *Abt SRBI, Inc.*

Dean Williams, *Abt SRBI, Inc.*

Amanda Geller, *Columbia University*

Tom Tyler, *Yale University*

With increasing budget constraints, it is progressively becoming more challenging to come up with defensible survey research design for targeting rare population in smaller geographic areas. Cost savings can be achieved by effective sample design and efficient household screening process. In designing such samples, there are methodological considerations we should be aware of. We present in this paper our sampling approach, an innovative geographic screening procedure, frame comparisons and address cost precision trade-off.

We present data from the New York City Stop and Frisk study conducted among a probability sample of 1,261 males age 18 to 26 in 37 New York City low SES neighborhoods. The challenge of screening for this age group is exacerbated by small area sampling. We present an innovative real-time geocoding based solution for targeting smaller neighborhoods for RDD samples and evaluate its effectiveness. In addition we examine hybrid Dual-frame RDD and List sampling approach for generating a probability sample of 18 to 26 year old males in small geographic areas. We explore methodological considerations including cost-precision trade offs, multi-frame weighting and frame comparisons.

## **Building Survey Organizations to Effectively Meet the Challenges of Measuring Public Opinion in Today's World and Tomorrow's**

### **The Unique and Not So Unique Position of Academic Survey Research Centers: Planning for Long-Term Sustainability, Quality, and Cost Excellence**

Stacey Giroux, *Indiana University*

Academic survey research centers occupy a unique space both within the field of survey research and on their respective college campuses. Striking a balance between providing services to others doing research, both within and outside of the university, in addition to producing original research, is an ongoing challenge and one that is often met without extensive university resources. In addition, university-based clients often have different needs from clients in the private sector. Yet, in some ways, academic survey research centers are not that different from small research firms and organizations. We face similar challenges that include increasing cost pressures and the need to adapt to new approaches that are being used to measure public opinion. We present data from a recently fielded survey of Indiana University faculty that gathered data about their research plans, needs, and awareness of research services offered by the Indiana University Center for Survey Research. This survey was used to inform the Center's strategic plan, and some of the results were quite unexpected. Other elements of the Center's strategic plan address the development of detailed workflows and quality checkpoints to improve quality, as well as the use of real-time metrics to drive survey and project decisions—ultimately, we are trying to learn how to do more with less. As part of this panel discussion, we share lessons we have learned in our strategic planning process and would be interested to hear about the experiences of other small research centers and organizations that may be addressing similar issues.

### **Rewards and Challenges of Establishing an Innovation Institute Within a Commercial Organization**

Michael W. Link, *The Nielsen Company*

Innovation truly is the life-blood of any research organization -- stay ahead of the curve or risk becoming irrelevant. The question is how to do this within an environment of rapidly changing technology, declining public participation, and shrinking budgets. In 2012, Nielsen established the Measurement Science Institute, a virtual organization that operates globally across Nielsen's 300+ methodologists, statisticians, and demographers. The Institute pursues goals in three basic areas: (1) methodological innovation, (2) scientific stature, and (3) academy company status. Some of the specific activities undertaken involve development of dedicated global research and development teams, establishment of on-going relationships with more than a dozen research universities across the globe, launching a senior Fellows program and a series of leadership programs to accelerate staff development. The rewards can be substantial, such as faster development (and sometimes fail) times for new projects, greater recognition of the complex science being undertaken at Nielsen, and a growing "pipeline" of top-level talent. The challenges in getting such a program off the ground, however, are not inconsequential. Obtaining "buy-in" from leaders who already have "full plates," organizing actions and programs when staff members are not centralized or co-located, and (as always) securing adequate funding. As part of the panel discussion, we will explore these and related topics, discussing the "lessons learned" within a context that is applicable to an array of research firms regardless of size or sector.

## **On a Road Less Traveled: Making a Difference in Market Research**

Paul Braun, *Braun Research, Inc.*

Braun Research has sustained its niche as a hybrid data collection company with experts in the social and political polling arena. We leave analytics to full-service firms and concentrate on what we do well: sampling, sample design, and execution. We have a rigid quality control process and ensure everything we do stands up to public and internal scrutiny.

All this has proved successful, but we work to continue to stand out. The ways we maintain success and differentiate include my focus for the panel:

- 1) Show how Internet research, Facebook and social media are changing data collection, sampling, recruiting, and execution. How Braun Research embraces and implements advances to foster successful and defensible survey outcomes; and
- 2) The lifeblood of our business is knowledge and to sharpen skills we will administer a monthly, internal online poll assessing our team's knowledge of current news. Expanding awareness of the world makes work smarter, and integrates that needed 'extra' when answering the important client questions. This is just underway. Our objective is to increase interviewer's knowledge and reduce refusal rates. Refusals appear tied to respondent's questions not responded to well or fully. Further, interviewers who are "news junkies" have better response rates and production.

The future is unlimited. We are always striving to make a difference in a limitless world of possibilities.

## **Changing the Way We Manage Data Collection without Changing the Data Being Collected**

Jamey Christy, *U.S. Census Bureau*

The US Census Bureau has undertaken a number of initiatives to streamline the efficiency and effectiveness of its field data collection activities. Many of these initiatives make use of emerging technologies, allowing the Bureau to incorporate innovative approaches to the management of its data collectors. The most visible change was a significant reorganization of the field structure – the first such undertaking in over 50 years. The Bureau reduced its field infrastructure from 12 offices to six in just 12 months, substantially reducing the "overhead" costs required to collect data. Less visible were significant changes in the oversight and supervision of the 6,000-plus data collectors, an expanded focus on data quality and an unprecedented use of real-time metrics to monitor the data stream during the transition. Despite a highly-organized planning effort, there were numerous challenges and disruptions. During the panel presentation, we will offer a high-level view of the overall process as a basis for discussion, including a look at the critical question – "did it work?"

## **Implementing the "Research Practices" Initiative at Pew Research Center**

Jon Cohen, *SurveyMonkey*

The Pew Research Center has launched a major "research practices" initiative to improve how we store, access, and share data, to diversify our data collection methods, and to improve collaboration and dissemination. Part of this effort is to change workflow at the Center to better leverage the great power of decades of historical survey data, and to layer in "big data" advancements around organic data. But part of the initiative is also changing the mechanics of

the operation, including data handling, dashboards, and editorial review. We will discuss the significant strategic planning and tactical decisions that surround this effort, as well as what it means for organizational culture and staff skill development.

## **Using Incentive in Distinct Populations**

### **Monetary Incentives and Response Rates in Household Surveys**

Andrew R. Caporaso, *Westat*

Andrew Mercer, *Westat*

David Cantor, *Westat*

Reanne Townsend, *Westat*

This meta-analysis aims to quantify the dose-response relationship between monetary incentives and response rates in household surveys. It serves to update and augment the existing meta-analyses on incentives by analyzing the latest experimental research, focusing specifically on general population household surveys, and including the three major data collection modes (mail, telephone and in-person) under the same analytic framework. Using hierarchical regression modeling and literature from the past 21 years, the analysis finds a strong, non-linear effect of incentives. Survey mode and incentive delivery timing (prepaid or promised) also play important roles in the effectiveness of incentives. Prepaid incentives offered in mail surveys had the largest per dollar impact on response. Incentive timing appears to play an important role in the effectiveness of incentives offered in telephone surveys but not in in-person surveys. Our model estimates a null effect of promised incentives in mail surveys; however, given the dearth of experiments testing this type of incentive, we are unable to draw firm conclusions regarding their effectiveness. Survey burden and survey year were both negatively correlated with response overall, however neither were found to significantly impact the dose-response relationship. Survey sponsorship was not found to affect either response rate or incentive effectiveness. The development and results of the model will be discussed and dose-response estimates specific to mode and incentive timing will be presented.

### **How to Open a Door with a Packet of Seeds and a Magnet: The Effect of Non-Monetary Incentives on In-Person Study Response Rates**

Andrew N. Williams, *Westat*

Ryan R. Hubbard, *Westat*

Respondents who are at home but refuse to answer the door when interviewers visit are the bane of in-person surveys. These individuals increase survey costs and waste resources, yet they result in nonresponse. This paper examines the relationship between incentives and avoiders compared to standard refusers by summarizing a series of experiments conducted to increase survey participation through mailing inexpensive non-monetary incentives to initial avoiders and refusers. While past research has examined the influence of incentives on response rate by comparing cash versus non-monetary incentives (Singer, 2002, Ryu et. al 2005, Toepoel 2012) and contingent versus non-contingent participation (Church 1993, Besky 2010, Chmura and Yancy 2012), this research examines the focused use of low-cost material incentives to encourage older sample persons to open their doors and complete in-person interviews.

This work is part of the Medicare Current Beneficiary Survey, a longitudinal CAPI study of Medicare beneficiaries. Approximately 16,000 beneficiaries are interviewed every four months.

Four case/control methods studies were conducted:

1. Mailing a handwritten card with or without including a packet of flower seeds to respondents who refused an initial participation request.
2. Mailing a handwritten card with or without including a packet of flower seeds to respondents who refused to answer their door after repeated interviewer visits.
3. Mailing a project refrigerator magnet with a handwritten card which either refers to the incentive or does not refer to the incentive to respondents who refused an initial participation request.
4. Mailing a project refrigerator magnet with a handwritten card which either refers to the incentive or does not to respondents who refused to answer their door after repeated interviewer visits.

One key finding indicates that mailing a \$1 packet of flower seeds to respondents who refused to open their door to interviewers increased subsequent participation by over 9 percent.

### **Cost-Effectiveness of Monetary Incentives in a Dual-Frame Mail Survey**

Sherman Edwards, *Westat*

William R. Andrews, *NOAA*

Michael Brick, *Westat*

Howard King, *Westat*

The National Marine Fisheries Service of the National Oceanic and Atmospheric Administration (NOAA Fisheries) has conducted a series of pilot tests to convert its survey of recreational saltwater fishing from a random-digit-dial (RDD) design to a dual frame design using address-based sampling and lists of licensed anglers in coastal states. The current pilot test uses a one-stage mail survey with no telephone follow-up, and has included experiments testing (1) different questionnaire approaches to reduce “avidity bias,” or the greater likelihood of anglers to respond to the survey than non-anglers, and (2) different levels of monetary incentive with the initial questionnaire mailing. Avidity bias is a significant concern, since the survey results in estimates of total recreational fishing trips and there is no external source for validation or calibration. As expected, the presence of an incentive increased response, and larger incentives had a larger effect, but incentives did not seem to increase response more for non-anglers than for anglers, using coastal proximity or sample frame as a proxy for avidity (Blair et al, 2012). This paper will examine the cost-effectiveness of the incentives, i.e., at what levels if any the incentive “pays for itself.” It will also examine the effect of incentives and varying incentive amounts on survey measures and on the likelihood of response.

### **Continuing Experiments on Non-Monetary Incentives in Physician Surveys**

Paul C. Beatty, *National Center for Health Statistics*

Eric Jamoom, *National Center for Health Statistics*

Janey Hsiao, *National Center for Health Statistics*

This paper will present latest findings from a continuing experimental program on non-monetary incentives in mail surveys of physicians. Data presented at the 2013 AAPOR Conference showed that including high quality pens with initial mailings boosted physician response rates in a longitudinal survey by about 4% (significant at the 0.01 level) compared to a control group that received no incentive. The topic of the survey was electronic health records (EHR), and response rates were only significantly higher for those who had adopted EHRs; however, all physicians who received a pen were likely to respond to the survey earlier than they had in the

prior year, which improved efficiency and reduced data collection costs. Data for this brief focus on two follow-up experiments. In the first, we investigated the potential benefit of sending pens at multiple mailings (at a first and third mailing) as opposed to sending pens at single mailings (first or third mailings) or not at all. Data from the experiment reveal not only how much incremental gain is achieved by sending multiple pens, but also the overall cost-effectiveness of the various conditions. In the second experiment, we explore the cumulative effect of the pen incentives across waves of a longitudinal survey. In some conditions, respondents received pens in successive waves of the survey; in others, respondents received the pen only in the earlier year, only in the latter year, or not at all. Through this design we explore whether the effect of the pen “wears off” over successive years, as well as whether it is worse to discontinue an incentive that respondents may expect to continue, or to never offer one at all.

## **Impact of a Differential Incentive in a Telephone Survey with a Low-Income Population**

Holly H. Matulewicz, *Mathematica Policy Research*

Carol Irvin, *Mathematica Policy Research*

In an era of declining response rates, it is challenging to connect with hard-to-reach populations, especially for a single-mode telephone survey. Although incentives have long been used to increase overall participation in surveys, little is known about the impact of differential incentives and the degree to which they motivate low-income populations to respond to a survey overall or within a specific time frame. This paper explores the impact of the incentive structure used in a telephone survey of low-income persons enrolled in a state-run health insurance plan, comparing results across two rounds of survey administration. In 2010, the 12-week survey offered a \$20 post-paid incentive for completing the 15-minute interview. The design included advance and reminder mailings, a toll-free call-in number, outbound telephone follow-up, and locating for those without working telephone numbers. In 2013, the survey was repeated with a new sample drawn from the same population, but the field period was truncated to six weeks. In an attempt to attain response rates similar to 2010, despite the shortened field period, the \$20 post-payment was enhanced with a \$10 differential (\$30 total) for those who called in and completed the interview during the first week, before outbound calling began. The paper compares weekly response rates of the two approaches and describes the impact of the larger-than-expected response to the first-week differential on both interviewer morale and production costs (e.g., locating, interviewing). Offering the differential incentive enabled interviewers to complete a high volume of interviews with willing respondents in the first week, which helped them quickly master the interview and increased the confidence with which they approached sample members in the subsequent outbound calls. Further, the differential motivated people to call in who might not have otherwise been reached (those for whom no working telephone number could be located).

## **Topics in Telephone Survey Samples**

### **All Dual Users are Not the Same: An Experiment Comparing RDD Landline Plus Cell vs. Listed Landline Plus Cell**

Kyley McGeeney, *Gallup, Inc.*

Manas Chattopadhyay, *Gallup, Inc.*

Decreasing response rates have increased costs for dual frame telephone studies using random-digit-dial (RDD) landline and RDD cell phone sample frames. In order to counteract this

trend Gutterbock et al (2011) suggested replacing the RDD landline frame with a listed landline frame in their article “Who needs RDD? Combining directory listings with cell phone exchanges for an alternative telephone sample frame.” Using a listed landline frame means users who have an unlisted landline plus a cell phone can only be included in the sample via the cell phone frame. However differences between unlisted dual users from the cell phone frame and those from the RDD landline frame may lead to coverage bias. A limited number of studies have explored this bias thus far.

In order to better understand the differences between respondents from various sample frames Gallup conducted an experiment in April 2013 to compare productivity and coverage of these frames. Groups one was an RDD 1+ list-assisted landline sample frame plus an RDD cell sample frame. Group 2 was a listed landline sample frame and an RDD cell phone frame. Gallup interviewed over 2000 respondents per sample source using the same questionnaire and sample management procedures.

Gallup found demographic differences between the two landline samples. However most notably dual telephone users with unlisted landline phones from the RDD landline sample were significantly different from the unlisted dual users from the cell phone sample on demographics and key metrics.

### **Within-Household Selection for Telephone Surveys: A Comparative Experiment of Eleven Different Selection Methods**

Jenny Marlar, *Gallup, Inc.*

Manas Chattopadhyay, *Gallup, Inc.*

Jeff Jones, *Gallup, Inc.*

Stephanie Kafka, *Gallup, Inc.*

Frauke Kreuter, *University of Maryland*

Numerous within-household selection methods have been tested since the advent of telephone interviewing. However few, if any, published studies have compared a wide range of selection techniques in the same experiment. In addition, limited studies have been conducted using the addition of a cell phone frame for analysis. Landline and cell phone frames are known to represent demographically different groups of respondents, and methods that may result in the most representative demographics in a landline frame may actually skew the results of the total study when combined with the cell phone frame. The present study tested 11 different within-household selection methods with approximately 11,000 landline respondents. A parallel cell phone sample was also collected.

Respondents were randomly assigned to one of 11 within household selection groups, and answered a short questionnaire about political attitudes and behaviors, and demographics. They were also asked to provide information about the age, gender and birthday of other adult members of their household. The selection methods tested included one probability based method, four quasi-probability methods, and six non-probability methods. Each method was evaluated on four criteria: demographic representation, response rates, accuracy, and ease of execution. The demographic representativeness of the sample resulting from each method was examined for the landline frame and then combined with results from a cell phone frame. Response rates and cooperation rates were also considered for each method. The characteristics of all household members were used to determine how often the household informant handed the phone to the correct person. Finally, ease of execution was also

considered, based on qualitative feedback from the interviewing team assigned to the experiment.

## **Where Am I Calling? New Telephony Technologies and Implications for Respondent Location in RDD Samples**

Hanna Popick, *Westat*

Eric Jodts, *Westat*

Jonathan Wivagg, *Westat*

In recent years the shift from landline to cell-phone usage and resulting concerns about sample demographics have received the majority of attention in RDD sample research. At the same time new technologies such as voice over internet protocol (VOIP), Google Voice, and landline-to-cell porting have begun to present a separate problem in landline RDD samples. This has been evidenced in new types of voicemail and telephone recordings that are not easily coded under the traditional landline protocols as well as landline respondents answering the phone in unexpected locations. Initial research into nationwide RDD samples conducted at Westat has given us a baseline for measuring the scope of the problem. Here we present the results of that analysis and expand upon it with new research into the telephony services used by RDD respondents and their impact on the locational accuracy of RDD samples. In addition, we evaluate how respondents describe these technologies and present potential implications to the issues of weighting by number of landlines in the household and defining households across the typical categories of cell only, cell mostly, landline mostly and landline only. We envision this as a baseline effort for charting trends in telephony technology prevalence and their impacts on RDD sampling and sample accuracy.

## **The Increasing Challenge of Cell Phones to Japanese RDD Surveys**

Nicholas E. Synodinos, *University of Hawaii*

Yasuyuki Saito, *The Asahi Shimbun*

Regular use of RDD surveys by the mass media in Japan started during the early years of the millennium. Besides concerns about the decreasing response rates, recent criticisms of Japanese RDD surveys center on their use of only landline samples.

Asahi Shimbun included a question in its mail surveys (starting with that of June/July 2008) to ascertain the number of persons with home access to a landline phone. Since that time, these surveys (with RR2 ranging from 71% to 79%) have been conducted 2 to 3 times annually with probability samples of Japanese adults (20+) drawn from the Voter Registration Lists. Starting in 2010, the surveys included an additional question about preferences of telephone usage when at home.

Those that indicated that they had home access to a landline telephone declined from 91% (June/July 2008) to 87% (May/June 2013) but remained between 86% to 88% in the last 6 surveys (end of 2011 to middle of 2013). Cell phone only (CPO) users increased from 15% (in Feb/March 2010) to 19% (in May/June 2013) and cell phone mostly (CPM) users increased from 27% to 34% during the same time periods. The latest survey found that among those with home access to a landline telephone, 10% were CPO and 39% were CPM users.

For the latest survey, home access to a landline telephone and preferred usage from home will be examined as a function of selected respondent characteristics (gender, age group, occupation, education, income, socioeconomic status, marital status, household size, housing



type, cabinet support, party identification, geographic region and degree of urbanization of place of residence). Similarities and differences with previous findings will be presented and the implications of the findings for the future practice of RDD surveys in Japan will be discussed.

## **Oversampling Minorities under Random Digit Dialing Vs. Address-Based Sampling**

Lauren K. Warren, *RTI International*

Bonnie Shook-Sa, *RTI International*

David Roe, *RTI International*

Brian Head, *RTI International*

Oversampling of ethnic and racial minorities is a technique often used in survey sampling. When analyses of data from minorities are of interest and if minorities are sampled using those proportions that naturally occur in the sampling frame, the number of completed interviews in desired subgroups may be smaller than needed to obtain the desired power for statistical analyses. Therefore, one must adjust the proportions for selecting minorities upward so that an adequate number of respondents are sampled and the desired sample sizes are reached.

The Aligning Forces for Quality: Assessment of Consumer Engagement (AF4Q) is a survey of chronically ill consumers of healthcare residing in targeted geographic markets, ranging in size from single counties to entire states. One of the main goals of the AF4Q health survey is to reduce racial and ethnic disparities in healthcare, making minority oversampling, specifically African Americans, Hispanics, and Asians, an important aspect of the surveys' success. In an effort to combat increasing challenges to efficiency and quality brought about by nonresponse, this survey underwent a transition between waves from random digit dialing (RDD) to address-based sampling (ABS). This transition presents a unique opportunity to examine the advantages and challenges of minority oversampling using these two typical methods of forming sampling frames, enabling the comparison of minority oversampling processes under each of these methods.

This presentation will discuss challenges related to this minority oversampling including identifying minorities to sample, targeting geographic areas, utilizing Census data, considering mode of communication (i.e. landline, wireless, and mail), meeting expected sampling yields, and optimizing the sample allocation to increase efficiencies. Additionally, minority oversampling elements from the RDD and ABS waves will be compared, where appropriate. The effectiveness of the minority oversampling techniques implemented will be discussed.