

AAPOR

AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH



Annual Conference



Asking Critical Questions: Toward a Sustainable Future for Public Opinion and Social Research

2013 Conference Abstracts

AAPOR

68th Annual Conference

May 16 – 19, 2013

Seaport Boston Hotel &
Seaport World Trade Center
Boston, Massachusetts

WAPOR

66th Annual Conference

May 14 – 16, 2013

Boston University, Photonic Center
Boston, Massachusetts

www.aapor.org

Thursday, May 16

1:30 p.m. – 3:00 p.m.

AAPOR Concurrent Session A

Innovations in Traditional Questionnaire Evaluation Methods

Getting Your Money's Worth! Targeting Resources to Make Cognitive Interviews Most Effective

Jaki McCarthy, *National Agricultural Statistics Service*

Cognitive interviewing has long been hailed as an effective technique to evaluate and improve survey questions. However, cognitive interviews are typically resource intensive and thus conducted on limited sets of questions and with limited sets of respondents. To be most effective, questions that are most likely to have adverse impacts on data quality should be targeted. In addition, respondents most likely to exhibit problems with these questions should likewise be selected for testing. One way to target a subset of questions is to use available information from previous data collections to identify questions with the greatest number of quality problems. For example, high edit or item imputation rates, greater numbers of requests for assistance answering these questions, etc. Once a subset of questions has been identified as good candidates for cognitive testing, respondents must also be selected. Again, information from existing data sets can be used to identify characteristics of respondents most likely to exhibit problems. Data mining techniques, such as classification trees, can be used to determine the type of respondents most likely to contribute to low quality responses. These criteria can be used to select respondents for cognitive interviews. In addition, knowing the pertinent characteristics of these respondents may also suggest useful probes that can be included in the cognitive interviews. Once questions have been revised based on the cognitive interviews, the same indicators of quality can be used to measure the improvement in data collection using the new questions. This approach has been employed in making revisions to questions on the Census of Agriculture; a case study provided will illustrate how this is an effective use of scarce testing resources.

Conducting Cognitive Interviews Over the Phone: Benefits and Challenges

Harmoni Noel, *American Institutes for Research*

Cognitive interviews are commonly used in the survey research world as a pre-testing method to test survey questions before they go into the field. They can also be used to test comprehension of other printed materials such as fact sheets or research summaries for clinicians on a variety of treatments or conditions. Cognitive interviews can show how respondents understand a question and identify potential sources of response error in resulting survey data. Typically, they are done face-to-face; however, some target populations such as clinicians or farmers are very difficult to reach in-person and other interview modes such as telephone interviewing may be more feasible and less costly. An additional benefit of doing cognitive interviews over the phone would be the ability to generate a nationally representative sample. To date, little research has examined the effectiveness of conducting cognitive interviews over the telephone, but others have explored other alternative methods for

conducting cognitive interviews such as self-administered Web surveys with promising success (Edgar 2012). Many researchers are facing budget and staff time constraints at the same time that respondents are becoming harder to contact in person and demand for larger samples have increased. Alternative approaches for conducting cognitive interviews may be the way of the future. This paper will present insights into the logistics, benefits and challenges related to conducting cognitive interviews over the phone. Data will be generated from interviewing people working on different projects that utilized phone cognitive interviews to learn more about their experiences with this method. Findings will be based on a qualitative analysis of themes identified across their experiences. For example, it will explore the challenges related to not having body language signals as indicators of affect or intention during the cognitive interview process. In addition, this paper will draw some comparisons to in-person interviews.

Self-Administered Cognitive Interviewing **Jennifer Edgar, U.S. Bureau of Labor Statistics**

Cognitive interviewing is traditionally an in-person pretesting method. The interaction between the interviewer and participant allows for in-depth probing allowing the researcher to use spontaneous probes designed to elicit explanations of the participant's response processes. Cognitive interviewing has been used to explore several stages and potential issues in the response process, including comprehension, retrieval, judgment and response. Past research has shown that it is possible that the goals of cognitive interviewing can be met using an unmoderated format, where participants "thinking aloud" respond to scripted probes without researcher intervention (Edgar, 2011). This approach was found to be promising, in terms of the quantity and quality of data collected as well as potential efficiencies, in terms of costs and time required to collect the data. This study builds on past work; comparing data collected using traditional cognitive interviewing techniques to data collected using unmoderated interviewing via the web. The quality of information collected in both modes is compared to determine if unmoderated cognitive interviewing can capture data equivalent to what was collected in the traditional lab setting. Specifically, respondent retrieval and comprehension are studied to see if both aspects of the response process can be understood using information collected online. The efficiency of the unmoderated method will also be evaluated, in terms of the costs and resources required to collect and analyze the data.

Using Web Ex to Conduct Usability Testing of an On-Line Survey Instrument **Kristin Stettler, U.S. Census Bureau**

Generally, usability interviewing is conducted in-person, to allow the researcher to observe the interaction between the participant and the instrument, and to conduct in-depth follow-up probing. The Census Bureau and the National Science Foundation conduct a bi-annual survey of state government R&D. Given a tight timeframe and respondents who are geographically scattered, it would have been difficult and costly to conduct an adequate number of usability tests in person in time for the survey to go into production on schedule. Therefore, we researched interactive on-line options using web-based conferencing software. Using Web Ex, we were able to conduct a portion of the usability interviews remotely, where the researchers in the Washington DC area observed and interacted one-on-one with on-line survey users in several locations throughout the U.S. This paper presents the reasoning behind our decision to choose Web Ex, the pros and cons of doing the interviews remotely on-line and suggestions for others who may be considering usability testing in this manner.

The Web Option in Multi-Mode Surveys

The Effects of Pushing Web in a Mixed-Mode Establishment Data Collection **Chris Ellis, *RTI International***

Mixed-mode data collection is increasingly becoming a standard in survey research methods, especially when inclusion of Web-based data collection is anticipated to increase data quality (de Leeuw 2005; Dillman 2000; Schaefer and Dillman 1998). However, offering the respondent the choice of mode can lead to unintended results, such as increased complexity or lower response rates (Medway and Fulton 2012). While “pushing” a particular mode (e.g. Web) may increase use, it risks lowering overall response rates (Mooney et al. 2012). Thus, there often exists a tension concerning if, when, and how to transition ongoing collections to a mixed-mode methodology when its origins are single-mode, such as paper form or questionnaire. The Deaths in Custody Reporting Program (DCRP), a data collection measuring inmate mortality began in 2000. Authorized by Congress and funded by the Bureau of Justice Statistics (BJS), the DCRP collects data on the circumstances surrounding deaths occurring in state prisons and local jails. It is the only national statistical collection that obtains comprehensive information about deaths in adult correctional facilities. RTI and BJS embedded a methodological experiment within the 2012 mailing to test the effects of concurrently offering multiple modes, but with a “push” of the Web option for some respondents. All agencies in the data collection were offered login credentials and information to utilize the Web option. A treatment of withholding paper forms provided in prior years was introduced, with a control group receiving paper forms. Assignment to treatment and control groups considered prior years’ mode selection. We will examine the results of the experiment – including timing, response rates, data quality measures, and variable costs – associated with the subgroups in the context of a longitudinal establishment study.

Internet Response for the Decennial Census – 2012 National Census Test **Courtney N. Reiser, *U.S. Census Bureau***

The Census Bureau has already committed to using the Internet as a primary response option in the 2020 Census. With this commitment in mind, the 2012 National Census Test (NCT) was developed to research the design and implementation of a secure, user-friendly online survey instrument. The primary goal of the NCT was to evaluate within-household coverage strategies for an electronic survey instrument. A secondary goal was to evaluate self-response rates of various mixed-mode contact strategies. This paper will focus on that secondary goal. Experimental contact strategies, which build off previous Census and American Community Survey research, utilize an Internet Push methodology with additional reminders, new motivational wording, and various timing strategies for the paper questionnaire mailing. Under the 2012 NCT Internet Push approach, households did not receive a paper questionnaire in the initial mailing but instead received an instruction card with information on how to provide responses online. Paper questionnaires were mailed to households who did not respond by a pre-determined date. This paper examines the proportion of Internet responses and overall self-response rates, including Internet, telephone, and mail responses for each of six experimental contact strategies.

Comparing the Effects of Mode Design on Response Rate, Representativeness, and Cost Per Complete in Mixed-Mode Surveys Conducted in New Jersey **Ryan Tully, *Princeton University*; Amy Lerman, *Princeton University***

Through a meta-analysis of recent split design surveys, Medway and Fulton (2012) find that mixed mode surveys “offering concurrent Web option in mail surveys results in a significant reduction in the response rate” (p. 10). In 2011, Princeton University fielded three consecutive surveys among residents of Princeton, NJ using Web-only, concurrent Web and mail, and sequential Web and mail mode options. These surveys utilized nearly identical survey instruments as well as similar contact strategies as outlined by Dillman, Smyth, and Christian (2009). In analyzing the data, we did not find a statistically significant difference in response rates among the Web-only mode option (AAPOR RR3 50.2%) and the concurrent Web and mail option (AAPOR RR3 47.7%). However, we did find that the use of the sequential Web and mail mode option had a statistically significant higher response rate (AAPOR RR3 57.0%) than the other mode options. Our study further analyzed the impact of mode design on the representativeness of the respondent pool, the probability of joining an online panel, and the overall cost per complete. Our results showed that the use of the sequential Web and mail mode option produced a more representative respondent pool than other mode options and greater participation rate in our online panel. Additionally, the study further found that the use of the sequential Web and mail mode design produced substantial savings in the cost per complete compared to the concurrent Web and mail mode design.

Changing to a Mixed-Mode Design: The Role of Mode in Respondents’ Decisions About Participation in the Fifth Wave of Understanding Society’s Innovation Panel

Debbie Collins, *NatCen Social Research*; Martin Mitchell, *NatCen Social Research*; Mari Toomes, *NatCen Social Research*

Understanding Society is a large panel survey, involving 100,000 individuals living in households in Great Britain. In 2012, for the first time, a sequential mixed mode approach was piloted, involving first Web and then face-to-face data collection for non-responders to the Web. The questionnaire was designed to collect equivalent data in both modes, using a single instrument. The pilot was undertaken with members of Understanding Society’s Innovation Panel (IP), who may have taken part in up to four previous waves of data collection, all involving face-to-face interviews. Panel members were randomly allocated to either a mixed mode or single mode data collection group, the latter involving only a face-to-face interview. This was done, in part, to assess the impact of adopting a sequential mixed mode design on response rates. While the Web response was higher than expected, a statistically significant difference in response rates between the two groups (mixed mode and single mode) was found, with the response rate for individuals being lower among the mixed mode group. Moreover fewer interviews were achieved with all members of the household in the mixed mode group than in the single mode group. To understand more about why these differences occurred we undertook qualitative follow up interviews with members of the mixed mode group to answer two specific questions.

- Why were respondents in the mixed mode sample group, who did not respond by Web, less likely to participate in a face-to-face interview than those in the single mode group?
- Why were members of households where one other person had completed by Web less willing to take part in the survey, in either mode?

This paper addresses these two questions, presenting findings from the qualitative research and discusses the implications for panel surveys planning to move to a mixed mode design.

Utilizing the Web in a Multi-Mode Survey **Lekha Venkataraman, *NORC at the University of Chicago***

While there has been increasing interest in Web based surveys, little research exists regarding how the Web fits into a multi-mode survey and what techniques can be used to increase Web participation. The presentation will focus on two populations surveyed for the National Survey of Early Care and Education (NSECE), the center based and home based providers (~12,000 cases). Both populations had a choice of completion modes, CAPI, CATI, or Web, yet the Web yielded significantly more completes than the other two modes. The NSECE utilized various mail, phone and email prompting strategies as well as incentive strategies which proved to have varying levels of success. In this paper we will investigate what led respondents and interviewers toward Web completion rather than other modes, as well as which prompting strategies were most likely to result in increased Web participation.

Issues in Landline and Cell Phone Dual Frame RDD Survey Design

Benefits of a Cell Only Sample for Oversampling Households with Children or Entire Sample **Marcus Berzofsky, *RTI International***

The Ohio Medicaid Assessment Study (OMAS) is a large dual frame study designed to develop key health and health care utilization metrics for families living in the state of Ohio. The OMAS oversamples families with children, African Americans, and Hispanics while also trying to achieve accurate county-level estimates. A dual frame (landline/cell) sample was selected, with 75% of the telephone numbers allocated to landline and 25% allocated to cell phone. The oversample of households with children was recruited from both frames while the oversample of minorities was from landline numbers only. However, in all cases the cell phone sample produced a higher proportion of the populations of interest. In this paper, we model the OMAS field experience to test the hypothesis that an all cell phone sample might produce similar quality at the same or lower cost than a dual frame design. We examine bias introduced from an all cell phone sample. We also examine the cost/quality trade-off for achieving the over-sample goals of families with children and race/ethnicity. The paper concludes by suggesting other cell phone/land line allocation strategies to achieve OMAS, and by extension other similar survey, goals.

Special Considerations for Weighting Local-Area Surveys **Mike Battaglia, *Battaglia Consulting Group, LLC***

Local-area surveys such as the New York City Community Health Survey (NYC CHS) and the Los Angeles County Health Survey (LACHS) produce estimates for adults residing in households in NYC and its five boroughs, and in Los Angeles County, respectively. Both surveys target specific sample sizes of adults in geographic subareas: 42 United Hospital Fund

(UHF) neighborhoods for NYC CHS and 8 Service Planning Areas for LACHS. A key aspect of the weighting methodology for local-area surveys is post stratification to population control totals, e.g., age, gender, race/ethnicity, education, marital status, and home ownership, etc. Obtaining up-to-date control totals for subareas can be challenging when available population data for subareas other than those which are of interest (e.g. ZIP Codes, Census Tracts, and Block Groups). We discuss the strengths and weaknesses of the sources for control totals (the Census Bureau Population Estimates Program, the 2010 Census, the American Community Survey (ACS) tabulation program, the ACS public-use micro data sample (PUMS), and describe the construction of subarea control totals for the NYC CHS and LACHS. We then evaluate the impact of including or excluding adults in non-residential housing such as college dormitories, prisons and nursing homes. For example, when weighting a Manhattan neighborhood that includes a university, not limiting the population control totals to adults living in households there result in 6,782 too many adults age 18-29 after weighting (38,250 instead of 31,468). The inclusion or exclusion of populations in group quarters should be considered when constructing demographic control totals, particularly for subarea weighting.

Best Weighting Approaches in Dual-frame Phone Survey with Multiple Domains of Interest

Jamie Ridenhour, *RTI International*

During the weighting process dual frame telephone surveys require a step to account for the fact that dual phone type users can be selected from either frame. There are several existing methods to achieve this. Which approach is best is often survey specific. We will look at the OMAS which is a study with multiple domains/outcomes of interest. To determine which approach was best for OMAS we computed the weights using four approaches: single frame estimation, 50% composite, optimal composite optimizing on minimizing the overall unequal weighting effect, and optimal composite optimizing on minimizing the design effect for past year's income. We present the impact to standard errors that each approach had on a range our estimates and discuss which approach we think is best for OMAS and other surveys like it.

Calculation of Response Rates for Dual-frame RDD Surveys

Robert Montgomery, *NORC at the University of Chicago*

Dual-frame surveys that combine landline and cell-phone samples have become the standard for telephone surveys. Although the surveys estimates and weights are calculated from both samples, response rates are usually reported separately. We start by considering the goal of producing a combined rate and how that may determine the appropriate method. We then examine different methods for calculating combined response rates and provide some guidance for when separate and combined rates are appropriate, as well as which method to use when combined rates are appropriate. We also explore different options depending on whether the cell-phone design is screening or take-all.

Address-based Sampling (ABS) as an Alternative RDD: A Test in California

Matt Jans, *UCLA*

Address-based sampling (ABS) from the USPS Delivery Sequence File (DSF) presents a sustainable method to overcome historical coverage decreases in landline random digit dial

(RDD) frames, and reduce costs relative to dual-frame cell/landline RDD samples. DSF Coverage tends to be better in urban areas than rural areas, yet apartments with multi-unit “drop points” and other living situations in which households are not clearly defined by a single mailing address can be challenges in urban areas. Data collection challenges occur in phone surveys like the California Health Interview Survey (CHIS) because respondents complete the survey in a mode other than the one by which they were contacted. We evaluate procedural, cost, and data quality implications of an ABS protocol in two communities in California (total n=7274 addresses sampled from the DSF). Communities were chosen based on population size/density and percentage of Spanish speakers. The mailing protocol included three ‘full-packet’ mailings with a reminder postcard between the first and second mailings. Each packet included single-page, one-sided 12-item screener questionnaire (in English and Spanish) that asked for basic health and demographic information, a phone number, and interview language preference (Spanish or English). A \$2 incentive, return envelope, and English and Spanish versions of the cover letter and FAQ were also included. Households providing a phone number were called to complete the standard CHIS telephone interview. Households not providing a phone number were called if one was matched to their address through public records. We compare ABS responses to CHIS RDD (cell and landline) responses in the same geographic areas. We also compare respondents who provided a phone number on the screener form and those for whom we used a matched phone number. We evaluate differences in key health statistics in addition to response rates and demographics of responding cases.

Minimizing Nonresponse Bias

Evaluation and Use of Commercial Data for Nonresponse Bias Adjustment

Andy Peytchev, *RTI International*

Response rates have been declining, posing a substantial threat to survey inference due to nonresponse bias in survey estimates. Concurrently, commercial vendors have been amassing data on individuals in the country. These data include not only demographic variables, but also substantive variables that can be similar to the key survey variables. These characteristics make these data potentially valuable for nonresponse adjustments, but their properties for this purpose remain unevaluated. Of critical importance are the rate at which these data can be matched to survey samples, the accuracy of these data, and their relevance in being informative about nonresponse bias. An additional hindrance is how they can be incorporated into adjustments due to the high expected missing data rate. Of critical importance is how to incorporate these data into nonresponse adjustments. We propose and evaluate the use of multiple imputation as a method that allows for missing auxiliary data and can offer highly efficient estimates when the auxiliary data are substantially correlated with the key survey estimates. We augmented a random-digit-dial telephone survey on tobacco use with data from Experian to evaluate: 1) the match rate of sample members with demographic data from Experian, 2) the match rate with substantive tobacco use variables from the commercial data, 3) the accuracy of these data for variables that are available in both survey and commercial data, 4) the impact of the use of these commercial data for nonresponse bias adjustment when compared to external benchmark estimates, and 5) the use of multiple imputation to provide efficient use of these data for estimates that are adjusted for nonresponse bias.

Interviewer Observations vs. Commercial Data: Which is Better for Nonresponse Bias Correction?

Jennifer Sinibaldi, *Institute for Employment Research (IAB)*; Mark Trappmann, *Institut für Arbeitsmarkt- und Berufsforschung (IAB)*; Frauke Kreuter, *University of Maryland JPSM & IAB*; Brady T. West, *University of Michigan Institute for Social Research*

Survey methodologists are searching for better paradata to use in nonresponse adjustment models, ultimately hoping to find variables that are highly correlated with both the outcome of interest and the propensity to respond. This analysis examines the performance of two data sources that can be used for nonresponse bias correction, interviewer observations and commercially available auxiliary data. The analysis will determine which data source is more predictive of the survey outcomes and is therefore, a better candidate for nonresponse adjustment models. The auxiliary data and paradata examined in this analysis are: 1. interviewer observations recorded for household income and receipt of unemployment benefits, and 2. commercial auxiliary data indicating household income and unemployment benefit. The survey data will provide a gold standard for both income and receipt of benefits. To answer the research question, separate models will be run for the observations and the auxiliary data, predicting the gold standard. The model fit will determine which data source shares more (accurate) information with the true value, making it better for adjustment. In addition to informing researchers wishing to improve their nonresponse adjustments, the results will benefit survey managers by providing guidance as to which type of data on which to spend the survey budget.

Assessing the Reliability of Unit Level Auxiliary Data in RDD Surveys: NHTSA Distracted Driving Survey

John Boyle, *ICF International*; Andy Weiss, *Abt SRBI*; Paul Schroeder, *Abt SRBI*; Mikelyn Meyers, *Abt SRBI*; Kristie Johnson, *NHTSA*

With declining response rates in population surveys, non-response analysis to evaluate survey bias becomes increasingly important. In essence, we need to compare the completed sample with sample units not completed in the survey. Hence, data from auxiliary data sources is necessary for evaluation of non-response bias.

Although exchange level data derived from the Census is available for all sample units in landline RDD surveys, its' usefulness is very limited. More useful unit level data such as age, education, income, race, ethnicity, household size, and housing tenure is also available, but only for some units. This information is obtained by matching sampled telephone numbers to other data sources including credit bureaus. Unfortunately, the reliability of this data source has not been well established.

This paper is based on the 2012 National Survey of Distracted Driving Attitudes and Behaviors conducted by Abt SRBI for the National Highway Traffic Safety Administration. The survey includes a total of 6,025 interviews, including 3,100 interviews from a national landline RDD sample, an oversample of 782 persons aged 16-34 from landline sample, and 2,143 interviews from a national cell phone sample. Matched records from the auxiliary data base were obtained for 49% of completed interviews and 54% for household contacts not yielding a completed interview. Although almost no auxiliary data is available for cellphone sample, matched records were found for 75% of the national landline sample.

A relatively high match rate for completes (77%) and non-completes (71%) in the landline sample, coupled with a relatively high rate of agreement between interview data and the auxiliary data on a range of key characteristics, suggests that auxiliary data may be useful in

correcting some non-response bias. Indeed, it may permit targeted follow-up efforts, in addition to sample weighting, to improve estimates.

Responsive Design for Web Panel Data Collection

Annamaria Bianchi, *University of Bergamo*; Silvia Biffignandi, *University of Bergamo*

Many surveys today are affected by high nonresponse. This can be a detriment to survey quality since nonresponse causes systematic error (bias) in the estimates. A related problem is the need of survey costs reduction. Given the decreasing trend in response rates and the corresponding increasing resources needed to achieve preset response rates, taking measures only at the estimation stage is no more sufficient to overcome these problems. Measures need to be taken also at the data collection stage. In this direction, different forms of responsive design have been proposed (Groves and Heeringa, 2006, Särndal, 2011). The purpose of this paper is to study responsive design in the framework of Web panel data collection. This method of data collection is increasingly widespread for general population opinion evaluation and it allows disposing of many variables on the participation process. We explore whether this amount of information could be exploited in the framework of responsive design. We evaluate as well whether this method improves the estimates in terms of bias reduction and assess the consequences on the variability of the estimates. The empirical application uses data from two on-going probability-based household panels: the PAADEL panel (Italian panel for the agro-food sector) and the LISS panel (Dutch panel managed by CentERdata, Tilburg University). Using these databases, we artificially reproduce a set of experimental responsive designs based on alternative interventions in the data collection. Results are analyzed in a comparative way to evaluate the impact of this approach on the final estimates. Bibliography: Groves, R.M., and Heeringa, S.G. (2006), Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A*, 169. Särndal, C.E. (2011), The 2010 Morris Hansen Lecture: Dealing with Survey Nonresponse in Data Collection, in *Estimation. Journal of Official Statistics*, 27, 1-21.

Comparative Ethnographic Evaluations of Enumeration Methods Across Race/Ethnic Groups in the 2010 Census Nonresponse Follow-up and Update Enumerate Operations

Laurie Schwede, *U.S. Census Bureau*; Rodney Terry, *U.S. Census Bureau*; Ryan King, *U.S. Census Bureau*; Mandi Martinez, *U.S. Census Bureau*

Why do minority undercounts persist over censuses, despite efforts to reduce them? We briefly review past coverage-related ethnographic studies then use a 2010 Census ethnographic evaluation with a records check to identify possible differences among race/ethnic groups in factors affecting enumeration methods and possible coverage error. This controlled-comparison evaluation was done in eight sites targeted to the major race/ethnic groups—American Indian, Alaska Native, Native Hawaiian and Other Pacific Islander, Asian, African American, non-Hispanic white, Hispanic, and a general site—in personal-visit 2010 Census Nonresponse Follow-up and Update Enumerate Operations. In the field sites, eight ethnographers observed and taped (when permitted) live census interviews, watched for cues of possible coverage error, and debriefed respondents to decide where to count persons. In the records check, we matched and compared rosters of ethnographer-observed housing units from 1) the observed standard interview and 2) the ethnographers' assessments to special 3) localized final Census Unedited File datasets to identify inconsistencies across records in where to count persons. We identify

qualitative themes crosscutting the ethnographic site reports. We present records check results and assess whether cases of inconsistencies among rosters and characteristics of affected persons and households differ by race, Hispanic, or household type. Some factors that affected enumeration methods and possibly coverage include: interviewer-respondent interactions, including question rewording; difficulty in gaining access to respondents; problems in canvassing and enumerating in rural areas without standard addresses; language issues; and cultural variations. We also reference selected results from the “Behavior Coding of the 2010 Nonresponse (NRFU) Interview Report” (Childs and Jurgenson 2011) that was based primarily on analysis of audiotapes collected by the ethnographers in this evaluation. We suggest improvements for enumeration and coverage and new research.

Cross-National/Cross-Cultural Survey Research— A Session Dedicated to Janet A. Harkness

Playing Soccer with an Accent: Variable Meanings and Analyst Bias Clifford Young, *IPSOS*; Darrell Bricker, *IPSOS*

The Total survey error paradigm delineates the many sources of error in surveys. Variable understanding across respondents lowers validity both within countries across individuals and across countries. Error can occur at the design stage, at data collection, and during analysis.

Trends in International Data Collection Quality Monitoring Beth-Ellen Pennell, *Institute for Social Research, University of Michigan*

Data collection across countries is especially important given the cross-national variation in languages, cultures, and structure. In addition, differences associated with data collection can compound differences and lead to artificial differences which are related to variability in reliability and validity rather than true substantive variation across countries. It is important both to optimize comparability by focusing on functional equivalence and to be sure that design are successfully carried out. Improved data collection quality monitoring can facilitate this goal.

Cross-Cultural Perspectives on Surveys of the U.S. Hispanic Population Trevor Tompson, *Associated Press NORC Center for Public Affairs Research*; Paul J. Lavrakas, *Independent Consultant*

As the 3MC perspective emphasizes, differences exist not only cross-nationally, but cross-culturally as well. The Hispanic population in the U.S. illustrates that point with much of this population being recent immigrants and with many having limited English proficiency. Steps for maximizing comparability between Hispanic and non-Hispanic population in the U.S. are discussed.

Interviewer Effects on Respondent Processing of Survey Questions, a Cross-cultural Analysis

Timothy Johnson, *University of Illinois at Chicago*

In interviewer administered survey, data collection is an interaction between interviewers and respondents. When these two participants are from different cultures, communication between them may be hampered and the risk of misunderstandings and more measurement error increases. Interviewer effects are always valuable to study and especially in cross-cultural surveys.

Monitoring Local and Regional Developments

Polling in the Midst of a Natural Disaster: The ABC News/*Washington Post* 2012 Election Tracking Poll and Hurricane Sandy

Gregory Holyk, *Langer Research Associates*; Damla Ergun, *Langer Research Associates*; Gary Langer, *Langer Research Associates*; Julie Phelan, *Langer Research Associates*; Seth Brohinsky, *Abt SRBI*

Hurricane Sandy made landfall the evening of Monday, Oct. 29, nine days in advance of the 2012 general election. Political pollsters faced two questions: one, whether or not it was possible to gather reliable regional and national estimates in the storm's aftermath, and two, whether or not it was appropriate to call people in the devastated areas of the Northeast. Judgments differed. The Gallup Organization decided to suspend its daily tracking poll, declaring that the hurricane "had compromised the ability of a national survey to provide a nationally representative assessment of the nation's voting population." We preferred, instead, to proceed, and to base our judgment on the data themselves. We polled the night of the hurricane, and, based on our ongoing assessment of data quality, we continued to poll in the days of its immediate aftermath and continuously up to Election Day. This paper presents a close look at how we approached interviewer sensitivity and the validity and reliability of the estimates obtained by our tracking poll in the midst of a major destabilizing event, and reports on lessons learned in the process. We examine post-hurricane daily call efficiency, break-offs and variability in estimates of the key demographics and attitudes nationally, in the Northeast region, and in the New England and Mid-Atlantic census divisions, compared with these measures in the 11 nights preceding the hurricane. We conclude not only that it was possible to poll during and after the hurricane in a sensitive and ethical manner, but that our polling produced valid and reliable national and regional estimates of attitudes and maintained an essential flow of information at a time when accurate polling was most in need and in demand.

Tweeting the Chicago Teachers Strike: Using Organic Twitter Data and Sentiment Analysis to Understand Support on a Local Issue

Nicholas D. Davis, *NORC at the University of Chicago*; Patrick van Kessel, *NORC at the University of Chicago*; Michael Jugovich, *NORC at the University of Chicago*

The September 2012 Chicago Teachers Union (CTU) strike and the response from Chicago Public Schools (CPS) were major media events during late summer and early fall 2012. With the rising popularity of Twitter, both the media and members of the public were able to tweet information and thoughts about the strike in great numbers. Our research examines tweets sent during the strike period to explore the use of organic data, as opposed to survey data or other experimentally designed data, for gauging public sentiment about the strike. Using the Twitter

Search application programming interface (API), NORC collected more than 125,000 strike-related tweets sent prior to and during the strike period. This presentation will focus on efforts to clean, deduplicate and process the collected tweets to facilitate their use in analyses of public perception on a substantive local issue. We employ natural language processing (NLP) and machine learning techniques for the purposes of conducting sentiment analysis. Using this information, we assess the relevance, sentiment (positive or negative tone), and position (for or against the strike) of the tweets and validate our processes using crowd-sourced manual coding. We conclude the presentation with a discussion of future research options and opportunities for the use of organic data in public opinion research.

From Red to Blue in the Green Mountain State: Real Change or Stability Against a Background of National Changes?

Richard L. Clark, *Castleton State College*; Ryan Flood, *Castleton College*; James McCormick, *Castleton College*

Prior to the 1992 presidential election, Vermont was traditionally a Republican state. From 1854 until 1963, Vermont's state government had been in Republican control, and Vermont was the most reliable supporter of Republican presidential candidates, favoring the Republican candidate in nearly every race since the inception of the Republican Party up until 1992 – with the sole exception of 1964, where Lyndon Johnson's landslide victory swept Vermont along in its wake. By most measures, Vermont was the most reliably Republican state in the union for a period of more than 100 years. Today, however, Vermont is perhaps the most reliable Democratic state in the union. It is the only state where the entire congressional delegation is comprised of representatives that caucus with Democrats (although Senator Bernie Sanders is nominally an independent) and the Democratic Party controls both the executive and legislative branches of state government. It is easy to mark the change from red to blue, with the historic election of Governor Philip Hoff in 1962 as the first Democrat in that position since 1854. Hoff's victory changed Vermont politics and set a path to competitive parties in Vermont. Despite the fact that we can identify when the change occurred, it has not been well established why the change occurred. Using public opinion data, Census data, and exit polls, this paper examines how Vermont became one of the most reliably Democratic states in presidential politics. Using those data sources, our paper tests the following two hypotheses: H1: Vermonters' political views have remained ideologically stable while the national parties have moved to the right. H2: Vermonters' have shifted their views away from the right over the past two generations, being aided by an influx of in-migration that has brought more liberal views to Vermont.

A Comparison of Live and Automated Congressional Race Pre-Election Polling **Meghann Crawford, *Siena College Research Institute*; Don Levy, *Siena College Research Institute*; Colin Frederickson, *Siena College Research Institute***

The Siena College Research Institute (SRI) has for three congressional election cycles accurately predicted many New York State swing congressional district races. Using live interviewers, SRI benchmarks the race in September and polls the district a final time within the last ten days before the election. A likely voter model is used in September and tightened in the final poll. In the recently completed 2012 election cycle, SRI simultaneously polled four New York State congressional races, all identified as among the top 75 most contested in the nation by National Journal's Hotline, in both September and late October using both live interviewers and interactive voice response (IVR) software. This paper compares the two sets of polls, live

and IVR at two time points, benchmarking in September and on election eve in late October/early November. In all cases, raw data is weighted by age, gender and stated party enrollment and only likely voters moved through the final screen. Regardless of any debate over weighting factors, both sets of data are weighted identically and compared not only to each other but also to the final results. We look at variation across the live and IVR by various demographics – party, age, gender – and across time points as well as the ultimate predictive efficiency of live as compared to IVR in these Congressional races.

The Growing Political Might of Ethnic Voters in California Elections

Mark DiCamillo, *Field Research Corporation*

According to exit polls Latinos, African-Americans and Asian-Americans comprised about 40% of California voters in the 2012 elections, a record high proportion. While the demographic changes taking place have been many years in the making, the 2012 elections may prove to be a turning point in California politics. My paper will trace the growth of ethnic voters as a share of the state registered voters. In addition, the paper will document the increasing tendency of California Latino and Asian-American voters to support Democratic candidates and will identify factors behind this change. The paper will draw primarily from the results of recent multi-ethnic Field Polls conducted in six languages and which over-sampled ethnic voter populations in seven of the ten statewide Field Poll surveys conducted in the 2010 and 2012 election years.

Reluctant Respondents and Data Quality

Using Doorstep Concerns Data to Study the Relationship Between Reluctance and Measurement Error

Ting Yan, *Institute for Social Research, University of Michigan*; Shirley Tsai, *U.S. Bureau of Labor Statistics*

Are reluctant respondents poor reporters? This is a question that the survey research field has been trying to answer for decades. Researchers have tried to answer this question from many different angles and the evidence is mixed. This paper approaches this question using doorstep concerns data. One type of paradata, doorstep concerns data 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 – Perceived Concerns Index (through principal component analysis) and Reluctance Class (via latent class analysis). We've found that the two measures are effective in characterizing and assessing the level of reluctance of survey respondents. In this paper, we will investigate the association between the level of reluctance exhibited by survey respondents and the quality of their responses to the survey questions making use of the two summary measures. We will attempt to provide further empirical results to the question: "Are reluctant respondents' poor reporters?"

Patterns of CATI Survey Break-off by Item Sensitivity and Respondent Characteristics

Ayesha De Mond, *Mathematica Policy Research*

Non-response and break-offs may bias survey findings. Theoretical frameworks for survey participation suggest the decision to initiate and complete a survey depends on the survey design and respondent characteristics, as well as psychological and social factors such as the cognitive demand of information sought, the sensitivity of items and the respondent's motivation and interest in completing the survey (Beatty & Herrmann, 2002; Peytchev, 2009).

Understanding determinants of response behaviors is particularly relevant for impact evaluation studies where differential non-response and break-off rates between treatment and control groups may compromise the validity of the study. However, literature on break-offs in the context of program evaluation is scarce. This paper will examine patterns of respondent break-off in the baseline surveys of the Parents and Children Together (PACT) Evaluation study. The PACT Evaluation consists of multiple components; here we focus on the experimental impact evaluation of a subset of Responsible Fatherhood (RF) and Healthy Marriage (HM) federal grantees undertaken by the Administration for Children and Families (ACF) with assistance from Mathematica Policy Research. The baseline surveys will gather descriptive information on study participants to make it possible to identify the characteristics of those who apply for RF and HM programs. The baseline survey instruments consist of 10 sections with questions tailored to respondents and rosters of family composition. The instruments collect data on sensitive topics such as relationship(s) with their child(ren) and partner(s), mental health, fidelity, economic stability, and experience with the justice system. We will examine the frequency of break-offs in relation to question content and respondent characteristics. We will explore break-off patterns and respondents' reasons for break-off through debriefings with interviewers. We will discuss findings and implications for survey design, response rates and data quality.

Nonresponse in Recontact Surveys

Besheer Mohamed, *Pew Research Center*; Greg Smith, *Pew Research Center*

One common way to identify individuals in hard-to-reach populations for surveys is to recontact respondents who indicated in previous studies that they are members of the population in question. For example, recontacting respondents who had identified themselves as Muslims, Asians or Mormons in surveys of the general public was one key component of the sample design for the Pew Research Center's surveys of these low incidence populations. But to what extent is nonresponse bias a problem in recontact samples of hard to reach populations? This paper employs logistic regression to compare non-response bias in re-contact samples to bias in samples acquired through random digit dialing. By analyzing non-response bias in recontact samples across three Pew Research Center surveys (including surveys of Muslim Americans, Asian Americans and Mormons), this new study extends and builds upon preliminary analysis presented at the 2011 AAPOR conference, which focused primarily on analysis of the Muslim American survey. The results will help researchers better understand both the advantages and the potential drawbacks in employing recontact sample as a means of surveying hard to reach populations.

Does Reissuing Unproductive Cases in a Face-to-Face Survey Reduce Nonresponse Bias? Evidence From the UK Citizenship Survey

John D'Souza, Ipsos MORI; Patten Smith, Ipsos MORI; Kathryn Gallop, Ipsos MORI; Angela Thompson, Ipsos MORI

It is common practice in UK face-to-face random probability surveys to reissue a subset of unproductive sample members to another interviewer in order to improve response rates. This practice is expensive to implement, both because interviewers are paid at higher rates for covering reissued cases and because interview productivity is considerably lower for reissued cases. In order to investigate the improvements in accuracy of reissuing cases, we analysed a variety of key survey variables in the 2009/2010 round of the UK Citizenship Survey. This survey collected data on a range of issues, including measurements of: attitudes towards community cohesion, behavioural changes caused by the economic downturn and frequency of civic participation activities. Measuring the non-response bias of a survey estimate is not usually possible. However, under the plausible assumption that the full sample is less biased than the first-issue sample, we are able to estimate the difference in bias between estimators based on the two samples. Bootstrapping yields confidence intervals for this difference. The results of our analysis show that the effects of reissuing were highly question-specific. For most variables, the estimates obtained from the first-issue sample were not significantly different from those obtained from the full sample. However, the differences were significant for many of the variables measuring frequency of civic participation activities. Furthermore, these differences could not be eliminated by non-response weighting. This implies that, for the variables measuring activity, reissuing does improve the accuracy of estimates. We discuss implications for existing survey practice and directions for future research.

Impacts of Unit Nonresponse in a Recontact Study of Youth

Jonathan Mendelson, Fors Marsh Group; Luciano Viera, Fors Marsh Group

When propensity to respond to a survey is correlated with key survey variables, nonresponse bias can occur. One method of assessing nonresponse bias is to compare respondents with nonrespondents using auxiliary variables from the drawn sample. A limitation of this method is that many frames have only basic demographic variables, which may be poorly correlated with response propensity. However, for low incidence and hard-to-reach populations, recontact studies are a popular option, often utilizing rich sampling frames containing behavioral and attitudinal variables from previous surveys. This paper assesses the impact of unit nonresponse in a recontact study of young adults who had recently completed a similar 'seed' study. Both studies were sponsored by the U.S. Department of Defense; the initial study examined attitudes and behaviors pertaining to military recruiting, and the recontact study assessed the awareness of and attitudes toward the Military's advertising campaigns. The seed study consisted of three iterations of a national mail survey of young adults ages 16 to 24, sampled from an address list database which covered more than 90% of the target population. Respondents to the seed study who provided an email address were used as a sampling frame for the recontact study, which was completed online. Using auxiliary variables from the original frame and from responses to the seed study, we examine unit nonresponse in the recontact study to assess differences between respondents and nonrespondents and the impact on key survey estimates. First, we compare characteristics of respondents and nonrespondents on a variety of demographic, attitudinal, and behavioral measures. Where characteristics differ significantly between the two groups, we conduct regression analysis to determine whether these

characteristics also significantly predict responses to survey questions in the recontact study. After examining the impact of unit nonresponse, we discuss implications for future research.

Methodological Briefs: Mode and Survey Error

Multi-Mode Survey Administration: Does Offering Multiple Modes at Once Depress Response Rates?

Jocelyn Newsome, *Westat*; Kerry Levin, *Westat*; Pat D. Brick, *Westat*; Patrick Langetieg, *Internal Revenue Service*; Melissa Vigil, *Internal Revenue Service*; Michael Sebastiani, *Internal Revenue Service*

As multi-mode surveys become the dominant methodology, questions have emerged about the optimal way to combine different modes. Is it best to offer all of the modes simultaneously, allowing respondents to choose their preferred mode of response, or is it best to offer first one mode and then another consecutively? Studies have shown that offering modes concurrently can depress response rates, a phenomenon sometimes called the “paradox of choice.” (Medway and Fulton 2012; Millar and Dillman 2011). According to this research, when respondents are provided with a choice of modes, they are less likely to respond by any mode. Consequently, there has been increased interest in determining how to best offer modes sequentially in order to increase survey response. For the 2010 IRS Individual Taxpayer Burden (ITB) Survey, an experiment compared a sequential administration (beginning with a Web survey) with a single mode, mail-only administration. The mail-only administration resulted in a higher response rate (44.1%) than an administration that offered first the Web survey and then the mail survey (40.9%). When planning for the 2011 ITB Survey, however, it was not an option to conduct a mail-only survey given federal government technology requirements. Therefore, it was decided that the 2011 ITB should follow the successful mail-only administration, with a simultaneous Web option. In an attempt to avoid the “paradox of choice,” the Web survey was offered in an understated way. While there has been very low Web survey response, overall response rates for the 2011 ITB Survey have so far been significantly higher than the 2010 survey (48.5%). This paper explores the success (and drawbacks) of this type of concurrent offering. The results of this administration suggest that it is possible to offer modes simultaneously if one mode is considered the primary mode and other modes are offered less prominently.

Tablets and Smartphones and Netbooks, Oh My! Effects of Device Type on Respondent Behavior

Hilary Ross, *Fors Marsh Group*; Jonathan Mendelson, *Fors Marsh Group*; Matthew Lackey, *Fors Marsh Group*

As the Internet becomes ever more accessible via smartphones, tablets, netbooks, and laptops, researchers have increasingly less control over how participants complete online surveys. Although options for online survey takers make these surveys more accessible than ever, researchers may not reap the benefits of increased accessibility if surveys are not configured to fit the wide range of devices available. Most current research on mode differences focuses on comparisons among paper, telephone, and online surveys, treating online surveys as a single mode. However, with so many devices available to access online surveys, researchers must consider the possibility that mode differences exist between devices within an online survey. This study examines respondent behaviors by device in a probability-based online advertising

tracking survey over a one-year period. The survey contains open- and closed-form questions with a variety of response option scales. Paradata from the survey administrator provides the browser user-agent tag, used to determine type of survey-taking device, and time to complete at the item level. This paper will examine the effect of the device on survey taking behaviors such as item nonresponse, open-ended response length, and straightlining. Implications for future online survey research will be discussed as well.

Reducing Survey Error in a Mobile Speech-IVR System

Michael Johnston, *AT&T Labs Research*; Patrick Ehlen, *AT&T Labs*; Fred Conrad, *University of Michigan*; Michael Schober, *The New School for Social Research*; Chris Antoun, *University of Michigan*; Stefanie Fail, *The New School for Social Research*; Andrew Hupp, *University of Michigan*; Lucas Vickers, *Parsons, The New School for Design*; Huiying Yan, *University of Michigan*; Chan Zhang, *University of Michigan*

Speech recognition systems for various automated tasks and transactions are now widely deployed. Despite advances, speech recognition is still not perfect, and designers of speech dialogue systems have various strategies for dealing with the imperfections. Can we live with this imperfection in speech-IVR survey interfaces? In principle, survey estimates should be accurate if misrecognition is unbiased—that is, if recognition errors are not systematic. We argue that the nature of the survey task should lead to different strategies for dealing with speech recognition error than other speech dialog tasks, which most often are initiated by the user. In a survey, adopting a high-accuracy dialog strategy with explicit response confirmation could frustrate respondents and increase break-off rates, while a low-accuracy-tolerant or no-confirmation strategy may be sufficient as long as the recognition errors are not systematic. In the current study we examine bias in recognition error in a corpus of 165 interviews on iPhones that ask numerical, categorical and yes/no questions, in a speech interviewing system designed specifically for the study. We compare a gold standard of human judges' interpretation of what respondents said to the speech dialog system's interpretation, to examine how spoken dialog system performance affects survey error for a range of different question types. Although recognition accuracy (agreement between human and automated judgments) was 94%, the question is whether the 6% recognition error was biased and in what direction. In particular, we examine the impact of dialog confirmation strategy on survey error and user satisfaction, and explore the use of acoustic and language model scores to limit errors. We also discuss which types of misrecognition were more likely, and what this suggests for the design of a survey instrument administered by a speech dialog system.

Mixed-Mode Data Collection in Health Care: Novel Approaches to Support Comparative Effectiveness Research

Margaret Good, *OptumInsight, Life Sciences*, Susan Brenneman, *OptumInsight, Life Sciences*

The American Recovery and Reinvestment Act (ARRA) of 2009 provided \$1.1 billion in funding to support comparative effectiveness research (CER). The intent of CER is to compare the relative effectiveness, benefits and harms of treatment options among different groups of patients in a “real world” setting. By improving our understanding of what treatments work best for whom and in what circumstances, CER helps physicians and patients make informed therapeutic choices. CER demands the development and expansion of a variety of data sources and methods. Optum has developed novel approaches to support CER by leveraging its

proprietary research database of administrative medical and pharmacy claims from a large U.S. managed care plan. This database allows for the identification of a targeted study sample and comprehensive analysis of health care utilization and costs, as well as treatment patterns, patient health outcomes and clinical characteristics. Limitations of administrative claims data are well known; in particular, the voice of the patient, the reasons for healthcare decisions and severity of illness defined by actual clinical lab values and vital signs are not available. In order to bridge the gaps in data gathered for reimbursement purposes, Optum engages in targeted primary data collection to obtain patient-reported outcomes via survey and clinical endpoints such as lab values via medical chart review. These data are combined with administrative claims data to explore a wide array of research questions, such as, the associations between treatment satisfaction, attitudes and beliefs about medicine and healthcare, and health status to treatment patterns and healthcare utilization; and association of severity of illness with healthcare utilization and costs. These designs provide an efficient and powerful methodology to conduct CER.

A Matter of Time: The Value and Optimal Timing of Follow-Up Questionnaire Mailings in a Multi-Mode Survey

Andrea Mayfield, *NORC at the University of Chicago*; Ashley Amaya, *NORC at the University of Chicago*; Kari Carris, *NORC at the University of Chicago*

Mail surveys remain popular in the United States primarily due to their lower costs relative to other interview-based methods of data collection. Inclusion of a mail component in a larger, multimode survey design may be used to increase response rates, obtain the requisite number of interviews, and contain survey costs. Dillman's Tailored Design Method provides a framework for the ideal frequency and timing of follow-up contacts to increase response rates in multimode surveys that include a mailed, self-administered questionnaire (SAQ) component. As the timing of mailings has not been tested recently, we seek to examine assumptions about the effectiveness, efficiency, and optimal timing of follow-up SAQ contacts in a survey of minority populations. We use data for this analysis from the Racial and Ethnic Approaches to Community Health Across the U.S. (REACH U.S.) survey, a multi-year project sponsored by the Centers for Disease Control and Prevention to eliminate health disparities among racial and ethnic minority populations. REACH U.S. uses a multimode, address-based survey design involving telephone, mail, and face-to-face interviews. In the latest round (Year 4), the REACH U.S. Survey incorporated a second SAQ mailing to non-respondents in all communities. The second SAQ mailing was sent six weeks after the initial mailing, in accordance with Dillman's Tailored Design Method. In our analysis, we find significant gains in the response rate by adding a second SAQ mailing. Additionally, we find that adding a second SAQ mailing is more cost efficient than additional contacts in other modes to achieve a target number of completed interviews. We also analyze the optimal time to mail a second SAQ mailing to achieve maximum response at minimum cost. Lastly, we investigate whether pursuing nonrespondents via multiple contacts changes key survey estimates and demographics.

Using Multiple Modes in Follow-Up Contacts in Random-Digit Dialing Surveys **Pranesh P. Chowdhury, *Centers for Disease Control and Prevention***

Recent studies have noted a decline in the response rates of random-digit-dialing (RDD) surveys. To increase participation and improve representation of the general population, the Behavioral Risk Factor Surveillance System (BRFSS) piloted several follow-up projects in 2012.

These projects included a mail follow-up study for landline phone numbers in 10 states (CT, KS, NH, IL, MA, MO, MT, ND, OH, and AR), a Web-based follow-up (WBFU) for landline phone numbers in 7 states (CT, DE, HI, IA, KY, NE and Washington DC) and a text invitation to Web follow-up for cell phone numbers in one state (CT). The purpose of the follow-up pilots was to test the feasibility of using landline/cell phone nonresponse contacts and their impact on demographic and health characteristics of respondents. All three pilots followed standardized protocols using specific non-responding RDD disposition codes to identify potential follow-up respondents. For landline follow-ups, phone numbers were matched to address and either entire surveys (for the mail follow-up) or letters with Web site links and login information (for WBFU) were sent to the household. Cell phone non-respondents with the same RDD disposition codes were texted and directed to the Web-site. Data collection will continue through December 2012. Results will be presented to illustrate unweighted differences in the characteristics of respondents of the three follow-up formats, as well as those who responded to the BRFSS. Preliminary data for the first nine months (N=1,107) from Web-based follow up survey indicate that it can increase the participation of female and Asian non-Hispanic respondents as well as those who have college degrees and annual household incomes of over \$75,000 or more. Single-adult households are also more likely to participate in the Web-based follow-up survey.

Where to Start: An Evaluation of Primary Data Collection Modes in an ABS Design
Ashley Amaya, *NORC at University of Chicago*; Felicia LeClere, *NORC at the University of Chicago*; Kari Carris, *NORC at the University of Chicago*; Youlian Liao, *Centers for Disease Control and Prevention*

As multimode address-based sampling becomes increasingly popular, researchers continue to refine data collection best practices. While much work has been conducted on Web + mail designs to maximize response rates, researchers have not yet tackled how phone + mail designs can be optimized. We use data from an experiment conducted on the Racial and Ethnic Approaches to Community Health Across the U.S. Risk Factor Survey (REACH U.S.) to evaluate two multimode case flow designs: (1) phone followed by mail (phone-first) and (2) mail followed by phone (mail-first). We use measures of response rates, cost, timeliness, and data quality to identify differences across case flow design. Because surveys often differ in terms of the rarity of the target population, we also examine whether changes in the eligibility rate alter the choice of optimal case flow. Results suggest that the mail-first design is superior to the phone-first design on most metrics. Mail-first achieves a higher yield rate at a lower cost with equivalent data quality compared to phone-first. While the phone-first design initially achieves more interviews compared to the mail-first design, over time, the mail-first design surpasses it and obtains the greatest number of interviews.

Thursday, May 16
1:30 p.m. – 3:00 p.m.
Poster Session 1

1. A Comparison Between Screen/Follow Item Format and Yes/No Item Format on a Multi-Mode Federal Survey

Sarah J. Hernandez, *NORC at the University of Chicago*; Svetlana N. Arakelyan, *NORC at the University of Chicago*; Vincent Welch, *NORC at the University of Chicago*

Over the last decade, methodological research (Dillman, 2008) has indicated that survey data quality can be increased if screener/follow questions (e.g., Do you have a disability? If yes, what type?) are replaced with yes/no questions (e.g., Do you have any of the following disabilities?). In keeping with this notion and consistent with government survey question format guidelines, the National Science Foundation's Survey of Earned Doctorates (SED)—an annual census of research doctorates awarded by U.S. institutions—changed the format of two demographic items (ethnicity and disability) from screener/follow to yes/no format. This work will explore the impact of this change on the responses to these items. To examine this effect, we will analyze the four most recent rounds of SED data (2008-2011); two rounds with screener/follow format and two rounds with yes/no format. Considering the previous research on this effect, we anticipate seeing higher levels of endorsement for both the presence of disabilities and ethnicity and fewer "other specify" responses on surveys with the yes/no format. We will concurrently explore whether the mode of administration (paper versus Web) moderates the effect of the format change. The SED is self-administered in paper and Web formats. When completed on the Web, the screener and follow-up items appear on different screens. While on paper they appear on the same page. Due to this difference, we anticipate that the effect of the format change will be greater for Web than for paper-and-pencil responses. The implications of these findings for survey design will be discussed.

2. Survey Weight Calibration With Multiple Imputation for Missing Data

Michael D. Larsen, *The George Washington University*; Benjamin M. Reist, *U.S. Census Bureau*

Multiple imputation (MI) fills-in missing information with two or more possible values. Observed data are used to model relationships among variables. Multiple draws for each missing value from conditional distributions enable representation of uncertainty. Calibration estimation in sample surveys adjusts survey weights so that estimated totals match control total targets. Post stratification and raking are versions of calibration commonly used in sample surveys and opinion polls. This paper examines calibration used in combination with MI for missing data. Performance of point estimators and variance estimators for estimated parameters are studied. The potential for calibration weighting with MI to reduce a source of bias in MI variance estimation is examined. Methods could apply to both sample survey and more general study design contexts.

3. Does Pre-Screening the Sample Improve Response in an Establishment Survey?
Julie A. Pacer, *Abt SRBI*; Kelly Daley, *Abt SRBI*; Marci Schalk, *Abt SRBI*; Jacob A. Klerman, *Abt Associates*

Establishment surveys are susceptible to a unique set of challenges compared to household surveys. Unlike a household survey, an establishment survey collects information from a representative of a business who speaks for that business, during business hours. While coverage may not be a problem in a survey of known establishments, nonresponse is a

considerable factor. The sampling frame may lack a contact name at the sampled business or it may provide outdated contact information. Furthermore, depending on the role of the intended respondent and size of the business, the intended respondent may not respond to a survey invitation due to competing work duties. Data from two recent establishment surveys allow the analysis of a strategy to improve response rates by improving the quality of the sampling frame. In both studies, a sample verification effort was performed pre-data collection to identify a particular respondent and for efficiency in main data collection. This research will explore the impact of the sample verification effort on response to the main survey by comparing outcomes such as level of effort, completion rate, and item nonresponse and including mode comparisons. The results will inform recommendations for future use of sample verification. Abt SRBI collaborated with Abt Associates on the Survey of Homelessness Prevention, sponsored by the U.S. Department of Housing and Urban Development, to understand the scope of the services being offered by agencies that received federal Homelessness Prevention and Rapid Rehousing Program funding. In addition, Abt SRBI with Abt Associates and the U.S. Department of Labor conducted a survey of employers regarding their use and understanding of the Family and Medical Leave Act that utilized the Dun & Bradstreet Market Identifiers file as a sampling frame. Sample verification was performed for each survey to verify the existence of businesses and identify a respondent.

4. Election Exit Poll Estimation Using Spatiotemporal Statistics
Clint W. Stevenson, *Edison Research*

There is an expansive amount of literature relating to Election Day forecasting during presidential elections. Most of the work on this topic relates to a national random sample and independent samples in key states. National samples provide insight on the nation as a whole. However, due to the way the Electoral College operates the state samples are critical to determine the winner of an election. This paper will examine the 2012 National Election Pool Exit Poll conducted by Edison Research on Election Day (November 6, 2012) and will take the spatial information into account. Geostatistical procedures are used to develop a spatial model of voting patterns using actual vote results as well as demographic and other information obtained only from the Election Day exit poll. Kriging and co-kriging is used to improve the spatial estimation of these voting patterns. These results from 2012 are compared to historical outcomes and exit poll data from the presidential elections in 2004 and 2008. The results presented here will allow for fine tuning attribute estimation both nationwide and at a state level.

5. Does Persistence in Nonresponse Follow-up Overcome Respondent Reluctance or Does it Contribute to Nonresponse?
Mary Frances E. Zelenak, *U.S. Census Bureau*; Brenna Matthews, *U.S. Census Bureau*; Mary C. Davis, *U.S. Census Bureau*; Jennifer G. Tancreto, *U.S. Census Bureau*

The American Community Survey (ACS) is an ongoing monthly survey that collects demographic, housing, and socio-economic data about people and households at approximately 3.54 million housing unit addresses in the United States each year. Since its inception, the ACS has collected data using three modes over a three-month period for each sample panel. In the first month, addresses are contacted by mail and households are asked to complete and return a paper questionnaire by mail. Beginning with the January 2013 panel, the mail contact will include instructions for an Internet response mode. Addresses that do not respond during the first month of data collection are contacted by telephone during the second month and data are collected using a Computer-Assisted Telephone Interview (CATI). Addresses that do not respond by the end of the second month

are subsampled and contacted using a Computer-Assisted Personal Interview (CAPI). During the mail contact month, households receive multiple mailing pieces, the number of which varies depending on whether and when a response is provided. Similarly, multiple contacts are possible during both the CATI and CAPI follow-up operations. Given that the CATI and CAPI operations are designed to target nonrespondents, it is very likely that some addresses are contacted numerous times throughout the three-month data collection period. These multiple contacts may lead potential respondents to be reluctant or even refuse to respond to the ACS. Paradata from the CATI and CAPI operations will be used to assess the success of the current CATI and CAPI procedures in obtaining cooperation from nonrespondents when reluctance is encountered. Recommendations for possible improvements to the current procedures and suggestions for future research will be provided.

6. One Drink or Two: Does Quantity Depicted in an Image Affect Web Survey Responses?

Nuttirudee Charoenruk, *University of Nebraska-Lincoln*; Mathew Stange, *University of Nebraska-Lincoln*

Researchers sometimes place images in Web surveys to motivate participation or to illustrate the meaning of a question, but studies indicate that presenting an image affects responses (e.g., Couper et al. 2007; Couper et al. 2004). To date, this research has investigated changes in responses based on the type of image presented; for example, how presenting an image of grocery shopping versus clothing shopping changes reports of shopping frequency (Couper et al. 2004). Our study extends this research by examining how quantity depicted in images affects respondent reports. Respondents will be randomly assigned to receive one of two Web surveys. One version will present pictures of one cigarette and one alcoholic beverage as illustrations for smoking and drinking behavior questions. The other version will present images with multiple cigarettes and glasses of alcoholic beverages for the same questions. We hypothesize that the respondents in the single cigarette and alcoholic beverage image condition will report consuming fewer cigarettes and alcoholic beverages compared to respondents in the condition in which many cigarettes and alcoholic beverages are depicted. Moreover, we hypothesize that the quantity depicted in the images will affect how respondents consider themselves as heavy versus light smokers and drinkers. Respondents may compare themselves to the quantity presented in the image to judge whether they smoke and drink heavily. In addition to analyzing differences in respondent reports, we will use eye-tracking data to analyze the time respondents spend looking at the image and the frequency with which they look between the images and the questions and response options to try to further understand how images and their content affect survey responses. We will conclude with implications for the use of images in Web surveys and Web survey design in general.

7. Geographic Accuracy of Cell-Phone RDD Sample Selected by Area Code Versus Wire Center

Xian Tao, *NORC at the University of Chicago*; Benjamin Skalland, *NORC at the University of Chicago*; David Yankey, *National Center for Immunization and Respiratory Diseases*; Jenny V. Jeyarajah, *National Center for Immunization and Respiratory Diseases*; Phil Smith, *National Center for Immunization and Respiratory Diseases*

The assignment of geographic location to cell-phone numbers at the time of sampling is often inaccurate. This inaccuracy can lead to increased cost and bias for area-specific telephone surveys and to increased variance for national telephone surveys with area stratification (Skalland and Khare, 2012). The assignment of cell-phone numbers to

geographic location can be done either based on the area code of the phone number or based on the location of the wire-center associated with the phone number. In this paper, we compare state and local-area geographic inaccuracy rates of cell-phone numbers assigned to geographic location based on the area code versus the wire center using data from the National Immunization Survey and the National Immunization Survey – Teen, dual-frame RDD surveys sponsored by the Centers for Disease Control and Prevention and fielded by NORC at the University of Chicago. In addition, we present estimates of demographic differences between respondents with accurate and inaccurate geographic assignment, first with the assignment based on the area code and then with the assignment based on the wire center.

8. Hola or Hello? A Priori Assignment of Interview Language Using Demographic Flags
Ying Li, NORC at the University of Chicago

The Racial and Ethnic Approaches to Community Health across the U.S. (REACH U.S.) Risk Factor Survey is a set of CDC-sponsored community surveys used to evaluate progress towards eliminating racial and ethnic health disparities. The REACH U.S. survey targets racial and ethnic minorities in specific geographic areas using an address-based sampling (ABS) approach and a mixed-mode data collection protocol involving telephone, mail and face-to-face interviews. Since REACH U.S. surveys racial and ethnic minority subpopulations, identifying and gaining cooperation from households in which the primary language is not English is vital, as many of these households may represent less educated, and recent immigrant populations that may be significantly different from their English-speaking counterparts. In Phase 3 of the REACH U.S. survey, NORC appended a set of vendor-provided race/ethnicity flags to the sample frame. This allowed us to test the quality of these flags, as well as assess the usefulness of this a priori information in making decisions about how to approach these households. An analysis of Phase 3 data revealed that these flags are relatively reliable. In Phase 4, the flags were used for a priori specialty-language interviewer assignment. Flagged cases in communities targeting non-English-speaking ethnic subgroups were assigned initially to an interviewer who speaks that language. In this paper, we will analyze the effectiveness of language-based a priori interviewer assignments. Comparing Phase 4 to Phase 3, we will examine whether the use of these race/ethnicity flags increased survey participation and/or reduced the time spent completing the survey. In addition, key performance measures from the REACH U.S. survey will be analyzed to examine potential interviewer effects or differences in the resulting sample introduced through the assignment of specialty language interviewers.

9. Evaluation of a Targeted Dual-Frame RDD Sample of Sub-State Populations
Amy Couzens, RTI International

Due to declining coverage of the landline RDD frames, researchers have become increasingly reliant on dual-frame (cell phone and landline) RDD designs to maintain complete coverage of the household population. A key challenge facing users of this approach is achieving geographically accurate coverage of state and sub-state areas through targeted cell phone sampling. The Aligning Forces for Quality (AF4Q) initiative works to increase the overall quality of health care, reduce racial and ethnic disparities, and provide models for national reform through the alignment of efforts to increase public reporting, consumer engagement and quality improvement within these communities. The AF4Q consumer survey seeks to evaluate the effectiveness of the AF4Q initiative and has a target population residing in fifteen markets across the United States, ranging in size from single counties to entire states. The overlapping dual-frame design is comprised of both RDD landline and targeted cell phone samples. This poster presents data describing the geographic accuracy of the cell phone samples in each market and how the accuracy varied

by market size and characteristics of the population. Based on our findings, we will make recommendations for future dual-frame RDD studies of small geographic areas.

10. Using Maximum-Difference Scaling to Assess Community Values about Local Water Resource Management

Tom Eiland, *CFM Strategic Communications*; Edward P. Johnson, *SSI*

As a suburban community transitions from an agriculture-based economy to an industrial-based economy through rapid population growth, some of the core values and attitudes towards the environment can change. Washington County, Oregon has gone through this change over the past 20 years. Since 1990, the county's population has grown by 250,000 people (+70%) with high-tech companies, such as Intel, becoming the largest employers. To try and meet its new consumers' needs, Clean Water Services (a local sewer and storm water service district) wanted to assess the values and priorities for water resource management in the new community makeup. Working with CFM Strategic Communications and SSI, the District developed an online panel of approximately 30,000 residents. A total of 1,398 residents participated in an online survey on the relative importance of eight different uses of water. Instead of using a typical Likert scale where people could answer that all uses of water were important, a Maximum-Difference Exercise was used to force respondents to make trade-offs between each use of natural resources. Relative utilities were then compared overall and by different geographic locations to determine the values of each potential water use. In particular, the Maximum-Difference technique was extremely well suited to make respondents choose between categories that they might otherwise take for granted. As a result, the District collected richer data and was able to make more informed decisions on how to best allocate resources in a way that best meets the needs of its changing user population.

11. Are We Asking the Right Questions? An Exploration Into Crowdsourcing Survey Questions

Bryan B. Rhodes, *RTI International*

Survey researchers use several methods to develop research questions and gather corresponding survey items, particularly for omnibus style surveys. Researchers may include established survey items, call on expert panels, or conduct focus group or interviews with particular populations. These methods, however, can be limiting in the number of voices that are represented. Researchers (or research topics) who may be less established in a field may not be included. This could mean important gaps in knowledge or new strands of research are left unconsidered for a survey. One way of incorporating a much wider range of viewpoints when developing a survey is crowdsourcing. Crowdsourcing is defined as "a novel method of online, distributed idea generation, problem-solving, and decision making that involves an open call to a large, often undefined network or community of people ('a crowd'), to provide either independent or collaborative contributions to solving a problem or performing a task" (Dalal et al., 2011). To further explore this possibility, RTI International hosted a crowdsourced "Research Challenge." The challenge put a call out to researchers across a range of fields to submit a short research brief and up to 10 survey items. The submissions were blindly reviewed by a group of survey experts, and ten winners were selected to have their survey items fielded. This presentation will give an overview of how the "Research Challenge" was conducted. In addition, based on submissions and a survey of participants the presentation will explore the types of researchers who entered (and won) the contest, as well as their motivations. The results of the "Research Challenge" show that important research questions and survey items can come from a broad spectrum of researchers, that might otherwise be overlooked.

12. The Cultural Life-Course of Attitudes Toward New Medical Technologies: A Case Study of Xenografts

Mariah D. Evans, *University of Nevada, Reno*; Jonathan Kelley, *International Survey Center*

How do people decide whether new medical technologies are good, neutral, or evil? We explore this question through a case study of attitudes toward a rapidly emerging biotechnology which potentially could save thousands of lives annually: xenografts (xenotransplantation). This involves taking a human patient's own cells, modifying them so they can grow into an organ, for example a heart, and implanting them in an animal fetus (often a pig), sacrificing the animal after the heart has grown large enough, and transplanting the heart into the patient. Data are from a nationally representative U.S. sample survey (N=2069) with reliable multiple-item measurement of key concepts, analyzed by structural equation methods. The results show that public attitudes are largely positive, with differences mainly reflecting cultural stances rather than social structure. Consistent with relational anchoring theories, attitudes are strongly shaped by views on conventional human-to-human transplants. They are also strongly influenced by scientific knowledge and by acceptance of a Darwinian worldview. Demographic and religious differences are few. Extrapolating from these findings, we propose a hypothesis about the cultural life course of new technologies.

13. The Effect of Incentive Offer Timing on Interview Completion Rates for the General Social Survey

Beth A. Fisher, *NORC at the University of Chicago*; Mike Buha, *NORC at the University of Chicago*

Much has been written about the use of incentives in survey research, types and timing of incentives, and how this affects interview participation rates. Prior research suggests that the use of incentives can improve response rates in most types of surveys. Further, Singer, van Hooyk, and Maher (1998) found that providing incentives does not reduce future survey participation if incentives are not subsequently provided in panel surveys. However, feedback from our field staff during the most recent round of the General Social Survey suggested otherwise, with interviewers stating that offering incentives, particularly to panel cases that had previously received incentives, was critical to completing interviews. The 2012 General Social Survey began offering incentives to its panel respondents within one month of the start of the field period, typically after an attempt to obtain the interview without an incentive had failed. Interviewing staff was authorized to offer a fifty dollar incentive and, upon refusal and with project staff approval, to increase the offer in increments of fifty dollars to a maximum of two-hundred dollars or whatever their prior round incentive had been. All incentives were conveyed in both direct communication and refusal letters that were mailed to the respondent. By the end of the field period, all remaining cases, regardless of prior round incentive amount, were offered a two-hundred dollar incentive if the field manager thought it was necessary. Our poster will address this whether or not non-random incentive payments appeared to make a difference in interview completion rates. We will add to this analysis by examining how timing, differentials in the incentive amount between offers, and amounts of incentives offered in previous rounds of the General Social Survey could have affected response rates for panel and address-based respondents.

14. Social Media Usage Among Young Adults: What, How and Why?

Caitlin Krulikowski, Fors Marsh Group; Katie Solook, Fors Marsh Group; Yalcin Acikgoz, Appalachian State University; Jennifer C. Romano Bergstrom, Fors Marsh Group; Shawn Bergman, Appalachian State University; Fors Marsh Group

Social media hosts a tremendous amount of data about individuals and as such, has emerged as a new way for data collection. However, utilizing this resource effectively requires in-depth knowledge about social media usage. Information about social media user demographics for various outlets is critical to the extent that a targeted approach for data collection is needed. While many studies exist that explain the number of people using social media and the various types of social media people use, few provide specific details of time spent or behaviors and interactions on social media by different groups of users (e.g., gender, age, race). In this study, we sought to examine how young adults (ages 16-24) use social media. Specifically, we explored how much time they spend using various social media compared to performing other activities (e.g., reading, playing sports), how much personal information they share on each, and what specifically they do on each social media (e.g., post, read). To study young adults' social media behavior, we created a 58-item pencil-and-paper survey. Question topics included (but are not limited to) Internet and Social Media Usage (e.g., general activities comparison, interaction on social media), Future Plans (using social media to get information about future plans), and Current Experiences (employment, education). 3,743 participants completed the probability-based survey. Data demonstrate that most young adults use social media as much or more than they talk on the phone, play sports, and more, and there are usage differences between sub-populations. In this talk, we will show the different ways young adults from different groups use various types of social media and the amount of personal information they share on each. There is applied value in exposing trends of young adults' social media behaviors for researchers interested in optimizing social media usage.

15. An Alternative Approach to Measuring and Describing Trust as a Complex Socio-Cultural Phenomenon

Anastasia Mirzoyants, InterMedia Survey Institute

This study suggests a statistical model, which singles out trust when accounting for a range of factors that influence interpersonal relationship. Prior empirical studies examined trust from a qualitative perspective: through the description of participants' beliefs, experiences and behaviors (Blomqvist, 1997). As a result, there is no definition of trust agreed upon by different academic disciplines because most existing theories identify trust by describing its attributes rather than measuring it directly (Bloomqvist). The researcher uses the Rasch analysis to design a quantitative instrument that can be used to measure trust. Earlier attempts to use the Rasch model in social research demonstrated that the Rasch analysis enables a creation of a rigorous measure useful for focused exploration of complex phenomena common in various socio-cultural environments (Fisher, 1991; Irwin & Irwin, 2005; Johnson et al., 1995). The proposed measure relies on two theories of trust: first, the study of Bryk and Schneider (2002), who describe trust as the mutual positive evaluation of the relations participants according to four components: respect, competence, regard for others, and integrity. The second theory is Lewis and Weigert's (1985) interpretation of trust as a tri-level phenomenon, which consists of cognitive, emotional, and behavioral components. After a series of instrument calibrations, the researcher added one more level to Lewis and Weigert's theory, loyalty, thus, creating a 4X4 matrix-type measure of trust. The alternative measure was tested in a pilot study, which demonstrated that the measure captures the overall structure of trust and can help detect the differences in trust due to participants' demographic and/or socio-cultural characteristics, especially in the environments characterized by the power asymmetry and insufficient sense of belonging.

16. The Effect of Cognitive Dissonance and Effort Justification on Recruitment into a Longitudinal Survey Study of Military Families

Hope McMaster, *Naval Health Research Center*; Kelly Jones, *Naval Health Research Center*

Background: Despite substantial improvements to survey design and implementation, there has been a general decline in health survey participation in recent decades. A frequently cited barrier to participating is the considerable time and effort it takes to complete a comprehensive survey of health-related issues. Is it possible that requiring considerable time and effort of survey responders actually helps in the recruitment of new survey respondents? In order to address this question, the theory of cognitive dissonance via effort justification was used as a framework for designing an experiment as part of enrolling military personnel in a large prospective health study. Methods: The study population consisted of a random sample of 598 Millennium Cohort Study participants randomly assigned to either participate in a low effort task (completing 5 pages of a health survey) or high effort task (completing 24 pages of a health survey), before requesting their spouse's contact information, so their spouse could be invited to take a health survey similar to the one they were taking. Agreeing or not agreeing to the request for their spouse's contact information was considered a proxy measure for the participant's attitude about the health survey. Logistic regression was performed to investigate the adjusted associations. Results: There were 494 (83%) members of the original sample that completed the survey. After adjusting, respondents engaging in the more effortful task (N=258, 31% referred) prior to the request were more likely to provide their spouse's contact information than those engaging in the low effort task (N=236, 3% referred) (adjusted odds ratio 15.6, 95% confidence interval: 7.2–33.9). Conclusion: These findings suggest that spending considerable time and effort completing a health survey may actually increase general regard for the survey, and thus increase the likelihood of agreeing to subsequent recruitment requests.

17. Can't They or Won't They Answer Our Questions? The Implications of Satisficing in Attrition Analysis

Veronica Roth, *The Pennsylvania State University*; David Johnson, *The Pennsylvania State University*

Longitudinal data collection offers researchers the chance to explore change over time and establish temporal order, a necessary assumption for multivariate analysis (Johnson, 1988). If attrition is non-random, the sample may yield biased estimates. Satisficing, which occurs when respondents do not fully process a question when giving a response, and may falsely increase reliability of answers to consistent, but not valid, responses. Respondents who satisfice may be less invested in a survey, or they may have lowered cognitive ability to answer demanding questions in the survey (Krosnic, 1991). Satisficing has been linked to non-response in cross-sectional research, due to lowered cognitive ability (Kaminska et al., 2010). Using the National Survey of Fertility Barriers (NSFB), I will conduct an analysis of how satisficing may be related to attrition. The NSFB is a nationally representative RDD telephone survey, initially conducted from 2004-2007 and included 4,792 women aged 25-45 years old, with a 3 year follow-up interview with 3,723 respondents (Johnson and White, 2009). Using recency and primacy effects, reliability scores, education and duration of survey, I will test hypotheses that attrition is related to satisficing due to both cognitive ability and lowered commitment to the survey. I will then discuss the implications of these findings in the context of both reducing attrition and detecting bias in the dataset. Johnson, David. 1988. "Panel Analysis in Family Studies." *Journal of Marriage and Family* 50(4): 949-955. Johnson, D. R. & L.K. White (2009). National Survey of Fertility Barriers [Computer File]. Population Research Institute [distributor]. The Pennsylvania State University. University

Park, P.A. Kaminska, O., A.L. McCutcheon & J. Billiet (2010) Satisficing Among Reluctant Respondents in a Cross-National Context. *Public Opinion Quarterly*, 74(5), 956-984. Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5(3), 213-236.

18. Inauthentic Respondent Behavior

Arianne Buckley, *Arbitron Inc.*; Will Waldron, *Arbitron, Inc.*

Arbitron developed an electronic Portable People Meter (PPM) that automatically detects audio exposure to encoded radio signals. A key feature of Arbitron's PPM is its capacity to measure the listening behavior of each household member by issuing each person his or her own personal meter. The meter also contains a motion detector that allows Arbitron to determine whether the meter was carried each day. Panelists receive instruction and coaching to carry their meter, and only their own meter, throughout the day and most panelists comply very well with these instructions. However, as a quality measure, Arbitron has developed methods to determine when panelists are non-compliant with these instructions so they can be coached and/or removed from the sample as indicated. This study takes a closer look at inauthentic respondent behavior within Arbitron and how it can relate to other survey researchers. The study will examine the prevalence of this behavior and the motivations behind it. The analysis will investigate any patterns seen among these non-compliers in order to help develop strategies for predicting and preventing this behavior before it occurs.

19. The Interpretation of Aerial Imagery as an Alternative to In-Field Listing for Address Frame Creation in Rural Environments: A Proposed Methodology With Empirical Results

Becki Curtis, *NORC at the University of Chicago*; Ned English, *NORC at the University of Chicago*

While it is now possible to use address lists derived from the United States Postal Service Delivery Sequence file (DSF or CDSF) in urban and suburban areas, rural areas without city-style delivery may still necessitate in-field listing for address frame creation. Due to the resource intensive nature of in-field listing, many studies are not able to proceed with frame construction in rural areas. In an effort to understand the cost-benefit of listing procedures in rural areas, we have outlined a methodology for using aerial imagery for creating and validating housing unit lists as an alternative to in-field listing in rural areas. In so doing we present empirical results from a comparison of aerial imagery to an in-person listing of a segment in north eastern Montana. The in-person listing for this study was part of the 2011-12 NORC National Frame listing. The aerial listing was completed blindly without use of the DSF in order to determine what percentage of housing units could be found remotely. A similar study was completed by Dreiling et al. as a part of the National Children's Study (NCS), finding that alternative listing procedures were less time-consuming and more cost-efficient than "on-site" listing methods while still allowing for the identification of a large percentage of housing units (2009). Likewise, we determined that the aerial listing found 85% of the housing units listed in-person, while in-person listing took four times as long to complete and cost ten times that of the aerial listing. Our preliminary results indicate that the use of aerial imagery may be a suitable alternative to in-field listing in certain rural environments. Dreiling K., Trushenski S., Kayongo-Male D., & Specker B. (2009). Comparing household listing techniques in a rural midwestern vanguard center of the National Children's Study. *Public Health Nursing*, 26(2), 192-201.

20. Sample Responsiveness to Tracking Efforts on the SIF WorkAdvance 18-Month Study
Christy Aroopala, *Decision Information Resources, Inc.*; Jo Anna Hunter, *MDRC*; Lee Robeson, *Survey Management Inc.*

The success of longitudinal studies is directly tied to the quality of the respondent contact information in the sample (Laurie et al., 1999). Updated contact information on respondents prior to study launch can save valuable time and resources during the study and helps reduce nonresponse. Longitudinal studies with hard-to-reach populations, those that are mobile and low-income, require special attention to tracking and cohort maintenance since their contact information changes frequently (Duncan & Kalton, 1987). Recent research has begun exploring successful strategies for increasing response rates to tracking efforts in longitudinal and multi-wave studies (McGonagle, Couper, & Schoeni, 2011). This paper evaluates the tracking efforts implemented in the 18-month pre-launch phase of the SIF WorkAdvance Study. All 3400 participants are recruited to the study in monthly cohorts and are randomly assigned to either program or control groups. Participants are then surveyed 18 months later to evaluate program effectiveness. During the 18-month pre-launch period, tracking mailings are sent to respondents requesting their updated contact information at 6, 9, 12, and 15 months post-random assignment. These tracking efforts are currently underway and will continue through September 2014. This paper evaluates sample responsiveness to date for these three types of mailings: (1) 6-month greeting card with a magnet, (2) 12-month letter with a perforated bottom to return via Business Reply, and (3) 9-month & 15-month emails. Sample responsiveness will be evaluated with measures on the number of voicemail messages, Business Reply returns, and email updates received from respondents in response to tracking efforts. We also explore whether age or gender impacts responsiveness to these different types of mailings to assess possibilities for targeted tracking efforts in future work.

21. A Balancing Act of Politics and Brands: A Look at Corporate Donations to Political Candidates and *the Impact on Attitudes of Corporations, Politicians, and Purchase Behavior*

Whitney O. Walther, *University of Minnesota*

This past summer Chick-fil-a found was the center of a political controversy after Dan Cathy, chief operating officers, made several public comments opposing same-sex marriage. As a result, many consumers either boycotted or buycotted the company's product depending on their own stance on the issue. Similarly, posts spread via Twitter and Facebook last spring urging individuals to stop shopping at Urban Outfitters and American Apparel (known for attracting young, liberal-minded customers) after it was discovered that the CEO of the companies, Richard Hayne, and his wife donated to the campaign of right-wing Republican Sen. Rick Santorum and Santorum's Political Action Committee. The current study uses balance theory (Heider, 1958) to investigate the interplay of attitudes between brands and politicians. Using a 3 (corporation supports/opposes/neutral) x 2 (favoring candidate/opposing candidate) (N = 210) design, this study explores the way in which individuals attempt to maintain a cognitively balanced relationship between themselves, corporations that donate to a political candidate, and politicians. It is hypothesized that the higher one's level of support for either a corporation or candidate will determine the way in which he or she shifts his or her attitude of the corporation and candidate after discovering donations were made. For example, if one has a high regard Urban Outfitters, his or her opinion of Rick Santorum might increase after finding out about the donations made by Urban Outfitters' CEO. Alternatively, one who highly opposes Rick Santorum may be less likely to purchase items from Urban Outfitters after hearing of the donation. Results suggest general confirmation of balance theory, in that individuals wish to maintain a balanced relationship between their attitudes of corporate donors and politicians. General attitudes

toward the corporation and candidate, as well as purchasing behaviors are investigated. Results have both theoretical and practical implications for political communication research.

22. Designing and Defending Surveys Used in Commercial Litigation

Melissa Pittaoulis, *NERA Economic Consulting*

Survey evidence has become increasingly important in commercial litigation, particularly in intellectual property disputes. In this paper, I discuss the challenges of conducting litigation surveys. One set of challenges is encountered at the design stage. These include drafting questionnaires on unfamiliar topics, choosing the appropriate data collection mode, creating any stimuli that need to be tested, and working on tight deadlines. The second set of challenges is encountered after the survey is completed and the report has been submitted to the court. Surveys that are proffered as evidence in legal cases are held to a particularly high level of scrutiny. In most cases, the survey will be critiqued by a survey researcher hired by the opposing side. In addition, judges have varying experience evaluating surveys, and some may be quite skeptical of their results. Thus, the survey researcher must be prepared to defend his or her methodological choices. Relying on the academic survey research literature is one of the most effective ways to do this. However, the researcher must be aware of the areas in which the legal precedent on best survey practices and current academic opinion diverge. One example of this divergence is on the use of the “don’t know option.”

23. Voter Interpretation of Large Numbers in Politics: A Comparison of Data Collected From In-Person Solicited Surveys and Mechanical Turk

Brian M. Guay, *University of Richmond*; David Landy, *University of Richmond*

This poster presents data collected in two replicated experiments; one using Amazon’s Mechanical Turk and the other using a more traditional in-person solicited survey technique. Mechanical Turk is becoming increasingly popular in the field of psychology and cognitive science, though it has been slower to gain popularity in the fields of political science and public opinion. We explore the slow growth of this trend, while providing an analysis of data collected in replicate experiments. The experiment run using both types of data collection methods explores the effectiveness of a number-line intervention on voters’ interpretation of large numbers, such as a million, billion and trillion. American voters are being increasingly confronted with numbers of such large magnitude in daily political discussion of the budget, deficit and debt. Previous research shows that individuals often incorrectly use number comprehension techniques to estimate the magnitude of these numbers. In this experiment, participants are asked to rate a series of political scenarios based on real political events involving large numbers and to then place similar numbers on a number line ranging from one thousand to one billion. The experimental group is then presented with a similar number line, but with one million placed at its proper location. Participants are asked to evaluate a second set of political situations and number lines, thus demonstrating the effect of the experimental group’s exposure to the intervention task. The data collected using in-person solicited surveys and Mechanical Turk are presented and analyzed.

24. How Representative are Google Consumer Surveys?: Results From an Analysis of a Google Consumer Survey Questions Relative National Level Benchmarks With Different Survey Modes and Samples Characteristics

Parvati Krishnamurty, *NORC at the University of Chicago*; Erin Tanenbaum, *NORC at the University of Chicago*; Michael Stern, *NORC at the University of Chicago*

The decrease in coverage for traditional random digit dialing (RDD) samples is well documented (e.g., Blumberg et al. 2011). This decline in landline connections, particularly for young people, make coverage especially problematic (Keeter et al. 2007). Although

mobile phones can be added to landline sample frames to increase coverage, this dual frame approach introduces new challenges, as they are more prone to nonsampling errors in comparison to RDD and in the United States incoming calls are often counted against the respondent's minutes (Brick et al. 2011). Non-probability Web-based supplements have been suggested as a means to reducing problems with RDD coverage and picking up cell-only households without respondent-side costs. However, three questions need to be answered. First, do we find more cell-only households among non-probability Web samples? Second, how do these Web-based results differ from national level random sample results? Third, how demographically different are these samples from mode varying probability samples? In this paper, we present an analysis of a series of Google Consumer Survey questions including home cell-phone usage and compare the results to those from three national-level random sample surveys, all of which were cited in the AAPOR's Cell Phone Task Force.

25. Enumerating Households via a Mail Questionnaire

Charles D. Harm, *Arbitron, Inc.*

Arbitron is moving toward an address based sampling (ABS) frame, in an effort to reach a greater proportion of U.S. households. Currently, a mail-based screener questionnaire is sent to an ABS sample household where a selected address cannot be matched to a landline phone number. If a respondent reports being cell phone only or cell phone mainly the household is added to a cell-phone frame and used to supplement a 2+ list assisted RDD sample. As part of the current screener questionnaire, respondents are asked to provide information on the demographic composition of their household (e.g., race, age, language). In order to maintain a representative sample, households are selected to participate in the Ratings based on their household characteristics. The current demographic questions are relatively simple. Moving forward, our goal is to collect more detailed demographic information from households. Enumerating households gives us the ability to further stratify our sample, and focus our efforts on only attempting to recruit households that have desired household characteristics. Two approaches to household age/gender enumeration will be tested. One method involves asking for the "presence of" household members that fall into a defined age/gender category. The other method involves asking for the specific number of household members that fall into a defined age/gender category. Households will be enumerated via a follow-up phone call to assess the accuracy of the enumeration data collected via the screener questionnaire. Which approach to household enumeration will provide more accurate data? This presentation will examine the impact of household enumeration on response rates, and whether data quality is influenced by household demographics.

26. Alternative Strategies for Linking Longitudinal Survey Data Aaron M. Pearson, **University of Michigan Survey Research Center**

**Ryan J. Yoder, *University of Michigan Survey Research Center*; Lisa S. Holland,
*University of Michigan Survey Research Center***

As respondent concerns about confidentiality and reluctance to provide identifying information increase, linking participant responses across multiple questionnaire administrations can present a challenge for social science researchers. One approach is to have participants create a self-generated identification code (SGIC) to link questionnaires. This code is comprised of a group of items that are well known to the participant, easily recalled from memory, and remain stable over time. In this presentation we describe a four-element SGIC linking strategy consisting of month of birth, day of birth, last initial, and last four digits of the social security number. We examine the effectiveness of the resulting code for linking participant questionnaires in two distinct components of a large military study. In

the first component we examine the utility and validity of the linking strategy for respondents who were asked to complete a questionnaire spanning two administrative sessions, separated by a day. A unique identification code was assigned to participants to serve as the primary link between sessions, allowing us to assess the effectiveness of the alternative SGIC link by calculating match rates across sessions. We also examine the incremental utility of each SGIC element to successfully match questionnaires. In the second component, the SGIC became the primary link. This time we were interested in demonstrating the utility of the SGIC as the only source of information to link participant's questionnaires over three time periods relative to their combat deployment. The first session was conducted prior to deployment, the second took place immediately upon return (approx 9 months) and the third issued three months after return. Again, we assess the effectiveness of the SGIC by examining match rates across sessions.

27. Investigating the Bias of Alternative Statistical Inference Methods in Sequential Mixed-Mode Surveys

Zeynep T. Suzer-Gurtekin, *ISR - University of Michigan - Program in Survey Methodology*; Steven G. Heeringa, *ISR - University of Michigan - Program in Survey Methodology*; Richard Valliant, *ISR - University of Michigan - Program in Survey Methodology*

Sequential mixed-mode surveys combine different data collection modes sequentially to reduce nonresponse bias under certain cost constraints. However, as a result of nonignorable mode effects nonrandom mixes of modes may yield unknown bias properties for population estimates such as means, proportions and totals. The assumption of ignorable mode effects governs the existing inference methods for sequential mixed-mode surveys. The objective of this paper is to describe and empirically evaluate the proposed multiple imputation estimation methods that account for both nonresponse and nonrandom mixtures of modes in a sequential mixed-mode survey. The American Community Survey (ACS) or the 1973 public-use Current Population Survey and Social Security Records Exact Match data will be used to conduct empirical and simulation evaluations. The focus of the empirical evaluations and simulations will be mean family income and health insurance coverage.

28. The Nature and Dynamics of Candidate Trait Impressions

Scott Clifford, *Duke Initiative on Survey Methodology*; Sunshine Hillygus, *Duke University*

Character trait perceptions are important predictors of vote choice, yet we know little about the formation, nature, and dynamics of candidates' trait images. Using the AP-Yahoo 12-wave panel survey, we trace the individual-level evolution of candidate image throughout the 2008 presidential election. We then compare evaluations of Obama into the 2010 and 2012 elections. Initial analysis finds substantial partisan polarization in relative trait ratings throughout the 2008 campaign, a trend which is especially driven by initially undecided voters. Yet, in spite of the polarization, we find evidence that voters also update their evaluations as they learn new information. Not only do candidates maintain distinct character strengths and weaknesses throughout the campaign, even among party supporters, but individuals make greater distinctions between trait dimensions for any given candidate. However, once an individual has settled on a preferred candidate, they show greater consistency in their trait evaluations of their preferred candidate, suggesting a process of motivated reasoning. Trait evaluations of Obama, however, change again in response to information learned during his first term in office. Finally, our analysis accounts for panel attrition in the analysis and considers the implications of attrition for the substantive conclusions.

29. On Factors Affecting the Accuracy of Congressional District Level Polls

Masahiko Aida, *Greenberg Quinlan Rosner Research*

When it comes to polling accuracy, political polls should be the easiest to evaluate; we can compare final survey estimates against the actual outcome after the Election Day. While it is relatively easy to evaluate the overall accuracy of each poll, it is very difficult to have a holistic understanding of the roles played by various features of the poll on its accuracy. Often each survey has a different mode (ex. IVR vs. live), different survey date, different treatment of missing data, different weighting scheme and different question wordings. These varied features will make an apple to apple comparison of surveys quite challenging. The author has a unique opportunity to shed some light on this situation as he has access to the micro-level data of many congressional district polls. Using micro-level data, the author can standardize certain features of polls (ex. Using exact same treatment of missing data, using identical target for weighting adjustment) and evaluate the effect of survey mode and the effect of timing (ex. Number of days prior to the election day). The author will use 473 micro-level congressional district level polls from 2008, 2010 and 2012 to evaluate the impact of above factors in the accuracy of the estimates.

30. Evaluating the Effect of Remote vs. In-Person Training Modes on Data Quality

A. Rupa Datta, *NORC at the University of Chicago*; Micah Sjoblom, *NORC at the University of Chicago*; Jill Connelly, *NORC at the University of Chicago*; Karen Veldman, *NORC at the University of Chicago*; Vicki Wilmer, *NORC at the University of Chicago*

The National Survey of Early Care and Education (NSECE) is an integrated set of surveys with households with young children, and institutions and individuals providing care for young children. This project employed a mixed mode data collection protocol and required several hundred interviewers, who were responsible for multiple tasks: locating, contacting, recruiting, and interviewing sampled households and sampled establishments and individuals providing early care and education. In order to prepare interviewers for the multiple challenges they would face in the field, a dynamic training effort was essential. Budget limitations and schedule constraints, however, made it possible to train only a portion of interviewers in person, so a group of more experienced interviewers were trained remotely. These two trainings featured similar content, but differed in significant ways, mainly in how the content was delivered. We will evaluate the relative effectiveness of these two training modes, including discussing differences in the trainings themselves, and comparing costs, retention on the project, and other operational factors. The majority of the effort will be to look at interviewer performance during the field period with a special focus on data quality issues. The project team developed a set of field interviewer performance metrics from different data sources to monitor and evaluate performance on a weekly basis. These metrics were constructed using both paradata (e.g., timing data, records of call attempts) and questionnaire data (e.g., item non-response rates, quality of verbatim responses) to create measures of interviewer efficiency and data quality across approximately 100,000 completed screeners and interviews. These metrics along with other non-survey-related interviewer characteristics (e.g., interviewer experience, measures of previous performance, gender, languages spoken) will form the foundation of our analysis of the training modes.

31. The Process of Turning Audit Trails From a CATI Survey Into Useful Data: Interviewer Behavior Paradata in the American Time Use Survey

Nicholas Ruther, *University of Nebraska – Lincoln*; Polly Phipps, *U.S. Bureau of Labor Statistics*; Robert Belli, *University of Nebraska – Lincoln*

In recent years, using paradata as a tool to improve survey methodology has grown markedly. Audit trails, i.e. supplementary output from a computer-assisted survey program, catalogue the actions taken during a survey interview, such as key strokes, time information, and edit warnings. (Couper 2000, Mockovak and Powers 2008, Dahlhamer 2004) By examining the audit trails, researchers can investigate problems and other issues within the survey interview in a systematic fashion. The American Time Use Survey (ATUS) is conducted using the Blaise CATI program, which produces an audit trail for each interview. This presentation discusses the process of taking an audit trail from its original state to a final assessable form, the problems that occurred and the solutions, and the information gleaned from the resulting data. Focusing on the time diary portion of the ATUS, audit trail text was imported to Microsoft Excel, parsed and tabulated, and subsequently the data were imported into a SAS statistical program. Many useful indicators for diagnostic analysis of the instrument and the interviewers were obtained. Counts of error edit warnings, how the interviewer interacted with them, and the lengths of time per interaction were able to be calculated and compared. In the cases examined, interviewers were much more likely to choose only one of the three options given by edit warnings when prompted. Verbatim activity entries in the time diary, or activities not assigned a pre-programmed code in the CATI instrument, were associated with greater use of durations (length of time entered in minutes) versus stop times (entering specific time of day) for information on time spent on a diary activity. This information was taken from a sample of 103 audit trails, indicating that greater sample sizes for future research should yield a wealth of new and more specific knowledge.

32. Air Pollution vs. Greenhouse Gasses. Government Should Limit the Amount? The Impact of Question Wording

Volker Huefken, *University of Duesseldorf, Institute of Social Sciences*

Does “air pollution” seem like a less serious problem than “greenhouse gasses”? Does effect three different measures the public support for limit the amount? In an experiment embedded in a German national CATI survey, adults were randomly assigned to rate the seriousness of “air pollution,” “greenhouse gasses,” or “greenhouse gasses that cause global warming” on the support that the German government should limit the amount. It will be shown whether there is greater difference between social insecurity, left-right ideology, and postmaterialism. Thus, word choice may sometimes affect public perceptions of the climate change seriousness of support for mitigation policies, but a single choice of terminology may not influence all people the same way, making strategic language choices difficult to implement.

33. Does It Really Make a Fracking Difference?

Robert K. Goidel, *Louisiana State University*; Michael Climek, *Louisiana State University*; Lina Brou, *Louisiana State University*

One of the great challenges of survey research involves understanding when and how citizens develop opinions on highly complex and technical issues. How do citizens who often lack basic information on the political system develop an opinion on something as complex as energy policy? Narrowing the scope even further, how do citizens form opinions on an issue like hydraulic fracturing, a controversial and poorly understood technique for drilling natural gas? Does the use of the word fracking, a term commonly used for hydraulic fracturing, affect public support for hydraulic fracturing? And, if so, in what direction? In this

paper, we utilize data from the 2012 and 2013 annual Louisiana Survey conducted by Louisiana State University's Public Policy Research Lab to consider whether slight shifts in question wording affect public perceptions of the safety of hydraulic fracturing and support for state government action to encourage drilling. Preliminary data from the 2012 Louisiana Survey indicate that the use of the term hydraulic fracturing or fracking increases the probability that respondents will say the method is unsafe and reduces support for state government action to encourage drilling.

34. Survey Research and Social Media Monitoring During the 2012 London Summer Olympics: A Case Study

Linda Lomelino, *Social Science Research Solutions*; Melissa Herrmann, *Social Science Research Solutions*; Susan Sherr, *Social Science Research Solutions*; Robyn Rapoport, *Social Science Research Solutions*

Social Science Research Solutions (SSRS) and Social Strategy1 (SS1) collaborated on a pilot study during the three-week period surrounding the 2012 Summer Olympics in order to address a crucial question: How can social media and traditional survey research work together to generate insights and quality data? With the increase in Internet access and social media usage, how can survey research and social media "listening" combine to add qualitative depth to research findings and increase the value of the data collected? The primary objective of the pilot study was to test the joint capability of traditional survey research and social media data collection to observe attitudes and opinions and to better understand the added value of integrating these two methods into a single research endeavor. The study involved collection of data regarding respondents' viewing of the 2012 London Summer Olympics and their attitudes about both the event itself and their experiences consuming Olympic media content. Data collection occurred simultaneously through random sample telephone omnibus surveys and Web monitoring from July 18 through August 19, 2012, a week prior to, during, and a week following the 2012 London Summer Olympics. Looking at all three time periods allowed the research team to examine whether intended viewership of the event differed significantly from consumers' actual viewing behavior. Similarly, collecting social media data during these three time periods provided a benchmark of the volume and nature of conversations prior to, during, and after the event. In addition to profiling the people who use social media and the specific social media outlets that are being used most frequently, this study also demonstrated that, while there are parallels between data collected through a random sample survey and social media monitoring, these data also diverge in ways that are both interesting and important to market researchers and academics.

35. Potential Impact of Modifying the Fielding Time of a Web-Based Survey

Herb M. Baum, *Data Recognition Corporation*; Anna Chandonnet, *Data Recognition Corporation*

As the field of survey research looks for a sustainable future, greater emphasis is being given to conducting Web-based surveys. However little is known about the pattern of response for these surveys. The question our presentation will address is whether, in a Web-based survey of a closed population, the percentage of respondents' providing a positive rating changes by the timing of when the person responds. The United States Patent and Trademark Office (USPTO), to improve the quality of their work and comply with the Government Accountability and Reporting Act (GPRA), conducts a Web-based survey of patent examiners twice a year. The survey is designed to gauge the satisfaction of the patent examiners with the internal and external factors that impact their ability to provide high-quality patent examinations. According to Dillman (2009) "The optimal timing sequence for Web surveys has not, we believe, been determined yet. Moreover the timing will depend

on the nature of the survey and the population being surveyed.” In practice, many Web surveys are fielded for two weeks with an initial invitation message followed by a reminder one week later. However, despite our wanting to adhere to that schedule, either of the following often occurs:

- The survey field period is shortened. For example, there is a meeting next week and we need to close the study early and present the results.
- The survey field period is extended. For example, you received a low response rate and feel that by keeping the study open longer you might increase it to more respectable level. We will explore how our results would differ with alternate Web survey field times.

This research is a continuation of work that was presented at a regional evaluation conference in New Jersey.

36. Looking for Solutions to America’s Energy Problems

Jennifer Benz, *Associated Press NORC Center for Public Affairs Research*; Matt Kozey, *NORC at the University of Chicago*; Trevor Tompson, *Associated Press—NORC Center for Public Affairs Research*

The U.S. public, politicians, policymakers, and experts alike agree that U.S. energy policy is an important issue for the country and one where government needs to be part of the solution. However, as demonstrated in the contentious exchanges on energy policy during the 2012 presidential debates, consensus on the causes of America’s energy problems and the appropriate policy solutions breaks down across a partisan divide. The Associated Press-NORC Center for Public Affairs research, with funding from the Joyce Foundation, conducted a nationally representative household survey with 1,008 adults on landline and cell phones to measure the general public’s opinions about key energy issues in the United States. Additionally, the survey assessed how the public understands, learns about, and acts upon energy issues. Using multivariate regression, we find that party identification is a stronger predictor of opinions on energy issues than demographic and socioeconomic characteristics. While individuals in both parties agree that energy issues are important at fairly equal rates, party identification appears to be the strongest influence on perceptions of the causes of and solutions to this country’s energy problems. As expected, this is especially clear when looking at the partisan differences on alternative energy sources and domestic drilling policies as causes of and solutions to the country’s energy issues. However, among the many partisan divisions, we do find similarities on key attitudes that have important policy implications. Mainly, we find that the energy industry and utility companies have the potential to be accepted and trusted actors in policy solutions. The public believes that the energy industry shares more responsibility for increasing energy saving in the U.S. than the government or individuals. Additionally, utility companies are the only source of energy savings information that reaches a majority of the public and is considered a trusted source across party lines.

37. The Effect of Cell Phones on Uninsured Rates: A Comparison of BRFSS and the Louisiana Health Insurance Survey Estimates

Ashley Kirzinger, *University of Illinois Springfield*; Stephen Barnes, *Louisiana State University*; Dek Terrell, *Louisiana State University*; Robert Goidel, *Louisiana State University*

In this paper, we investigate how the inclusion of cell-phones in statewide samples affect estimates of uninsured rates in Louisiana. We utilize data from the 2011 Behavioral Risk Factor Surveillance System Survey (BRFSS) and from the 2011 Louisiana Health Insurance Survey (LHIS), a 10,000 household survey designed to estimate the number of uninsured children and adults. Both surveys significantly increased the number of cell-phone

respondents in 2011. In the BRFSS data, the uninsured rate for adults, 18 to 64, increased from 24.5 percent to 26.8 percent. In the LHIS, uninsured rates for children decreased from 5 percent to 3.5 percent while uninsured rates for adults increased from 20.1 percent to 22.7 percent. The CDC strongly cautions against describing these shifts as trends given the change in methodology. With this in mind, we ask a slightly different question: What would the uninsured rates have been without the inclusion of the cell-phone sample?

38. Effects of Response Format on Measurement of Readership

Randall K. Thomas, *GfK Custom Research, LLC*; Curtiss Cobb, *GfK Custom Research, LLC*; Julian Baim, *GfK-MRI*; Risa Becker, *GfK-MRI*

Estimating exposure to media sources, such as magazine readership, is a critical function that determines advertising prices. One method to determine the extent of magazine readership is through self-report, and can be done in a variety of ways, employing paper-pencil self-administered instruments, human interviewers using show cards, and Web-based surveys. When presenting a series of targets like magazines that serve as a filter for subsequent follow-up questions, there are a number of techniques that have been used, including a multiple response format ('Select all'), a yes-no grid (requiring a yes or no to each element), or a card sort task that separates the magazines into piles of 'yes' or 'no'. As part of an investigation to transition a magazine readership survey to a Web-based mode, we experimentally investigated alternative response formats to determine readership in the past 6 months. We converted the traditional human interviewer card sort task into a drag and drop task whereby magazine titles would be displayed in a single pile and respondents would drag and drop the magazines into 3 piles – Yes, read; Not sure; No, did not read. The Yes-No grid also included a middle category 'Not sure'. The multiple response format presented magazines with 4 in a row with 4 columns, and a response at the bottom 'I did not read any'. Each format had multiple screens to accommodate over 250 magazine titles. We found that the drop-and-drag format took the longest to complete, while the multiple response format took the least amount of time. In addition, the drop-and drag format show a 50% higher readership rate over the yes-no grid, and the yes-no grid showed a higher readership than the multiple response format. We compare our results to those found with high quality in-person interviews on readership.

39. The New Era of Innovative Incentive Treatments: Efficacy of Grand Prize Sweepstakes versus Costly Individual Incentives

Ekua Kendall, *Arbitron, Inc.*

Arbitron developed an electronic meter that automatically detects audio exposure to encoded radio and TV and other media signals. Panelists are asked to wear their meter every day from the time they wake up to the time they go to sleep in order to measure their full media exposure. The meter has a motion detector that allows Arbitron to determine whether a panelist carried their meter on any given day and the panelist receives monthly incentives based on their motion data. There is seasonal variance in meter carrying behavior--with panelists less likely to carry their meter during times when there is likely variance in their normal daily routine, such as during holiday periods and during the Summer. Increasing individual incentives during these time periods are very costly. Over the last 3 years Arbitron has analyzed the efficacy of implementing a grand prize sweepstakes in place of individualized cash incentives for these seasonal periods. Previously when the study was in its infancy, we presented an AAPOR poster that was very well received with numerous post-conference follow up from attendees. Now going into in year-three of implementation there is greater data for continued analysis on the effectiveness of a sweepstakes incentive in a panel setting. There were also some lessons learned in the running of multiple sweepstakes on a yearly basis. This presentation will reveal performance

related metrics of diverse demographic groups and additional sweepstakes methods of varying prize money amounts and visual promotional materials. This presentation will also reveal other interesting findings and represents an expansion of our knowledge base in this area of alternative incentives that anyone interested in this promising area of study will not want to miss.

40. Analyzing American Trust and Confidence Utilizing a Mixed Mode ABS Nationwide Survey

Danna Moore, *Social and Economic Sciences Research Center*; Donald Beck, *Booz Allen Hamilton*; Bruce Austin, *Social and Economic Sciences Research Center, Washington State University*; Dave Schultz, *Social and Economic Sciences Research Center, Washington State University*

An important performance indicator for government is trust and confidence of the American people. Financial and health care services are both fundamental to the health and well-being of most Americans and families. The U.S. population has experienced tumultuous circumstances with the high rates of foreclosures, bank closures and fraud, high financial service fees, high health care costs and a shifting health care system. There is much interest in the performance of the health care and financial services sectors as related to consumer satisfaction and ensuing trust and confidence. This nationwide survey evaluating trust and confidence in financial services, health care services, and important American institutions provides a unique opportunity to perform analyses of mixed mode survey results. This study evaluates the impacts of weighting and nonresponse adjustments for an address-based sample frame survey. We explore the impacts of weighting and compare these results across survey modes on key survey measures.

**Thursday, May 16, 1:30 p.m. – 3:00 p.m.
AAPOR Demonstration Session #1**

PHIT for Duty: Exploring a Mobile Data Collection Framework

Stacey Weger, *RTI International*; Paul Kizakevich, *RTI International*; Randy Eckhoff, *RTI International*; Yuying Zhang, *RTI International*; Jennifer Lyden, *RTI International*; Vesselina Bakalov, *RTI International*; Stephanie Bryant, *RTI International*

PHIT for Duty is an applied research program, developed on behalf of the U.S. Department of Defense, for prevention of chronic psychological health issues and post-traumatic stress disorder (PTSD) among troops recently returned from deployment. The Personal Health Intervention Tool (PHIT™) is an innovative field-deployable self-help system that is intended to be used for secondary prevention of psychological health problems with early intervention of PTS symptoms and risk coping behaviors. The goal is to reduce the short-term impact of traumatic and operational stress exposures, reduce incidence and duration of stress-related health problems, improve quality of life, and reduce the risks for PTSD and other long-term stress-related injuries. The PHIT platform combines a smartphone or tablet and optional, nonintrusive physiological sensors. PHIT for Duty integrates a suite of health assessments with an intelligent executive program that recommends, tailors, and presents advisories based on established rules and processes. PHIT provides for collecting information (instruments), executing application logic (virtual advisor), and displaying output information (activities). It integrates data ranging from questionnaires to diaries to Bluetooth-linked sensors, including wireless heart rate, sleep state, and actigraphy sensors. The built-in logic processor executes

custom logic to change the behavior of the application and display custom output using different forms of media. The PHIT framework is flexibly designed to collect data from different sources, have runtime intelligence for dynamic analysis, be customizable, work offline, and run on multiple mobile devices. Data is stored locally in an encrypted database and uploaded periodically to a project server whenever wifi is available. User privacy is maintained via multiple layers and safeguards. This demonstration will provide an overview of the PHIT platform and PTSD intervention app, discuss lessons learned from beta tests, and demonstrate examples of unique, new survey capabilities using this mobile data collection platform.

Tablets as Data Entry Interfaces – Solving Data Cleaning and Transcription Issues During Data Collection

Michael Costello, *RTI International*

Due to the increasing complexity of survey work, tablets can provide a strong support system for data enumerators during collection. Software can be written to assist in reminding enumerators when to skip questions, what kinds of prompts are acceptable to use, or when to abort a survey due to responses provided. It can also ensure that crucial questions are not accidentally skipped during collection. For survey administrators, the benefits are even more far reaching: Instant access to data, metrics on enumerator pacing, instant data entry with no additional wait time, GPS mapping of dwelling and more. Tangerine, an open source data entry interface developed by RTI International, is the first tablet based data collection software custom-created to record student responses on early grade reading or mathematics assessments, yet flexible enough to capture common survey formats in a range of languages and scripts without requiring programming expertise. Surveys can be collaboratively designed using a simple Web-based tool, the Tangerine wizard, similar to Survey Monkey or Google Forms. Tangerine does not require connectivity during data collection to be usable in low-resource and low-bandwidth environments. At the same time, where connectivity, e.g. via mobile networks is available, the software allows for regular back-up of the data to the central server which in turn allows for immediate review and monitoring of data collection progress.

Designing Surveys for Tablets and Smartphones

Sabin Lakhe, *U.S. Census Bureau*; Elizabeth Nichols, *U.S. Census Bureau*; Murrey G. Olmsted, *RTI International*; Tiffany King, *RTI International*

Designing surveys for mobile data collection raises a number of programming issues for researchers. For example--should we design an app that can be downloaded by the user or a Web application that would be rendered based on type of a mobile device? What are the benefits or drawbacks using app or Web based surveys? How should we display questions to accommodate different screen sizes and formats of devices? Should the user use the keyboard to enter date or should we use date picker? Should we automatically capitalize the first letter entered for names and addresses? In preparation for the 2020 Census, the Census Bureau developed and cognitively tested several android apps for three different decennial census forms: a household level questionnaire; a group quarters questionnaire; and a questionnaire used for people who did not receive the household questionnaire. For this study, we used Android-based tablet and smartphone devices to conduct the cognitive and usability testing. The presentation will review the challenges faced by the project team, our findings from two rounds of interviews, and the design changes that we made as a result of testing. The audience will have an opportunity to use and compare the questionnaires designed for the tablet and smartphone and compare the differences themselves. We will also talk about next steps in this

work and how we plan to address some of the challenges of preparing these apps for use in the field.

Thursday, May 16, 1:30 p.m. – 3:00 p.m.
AAPOR Concurrent Session B

Factors Related to Survey Participation

Social Isolation and Survey Nonresponse: An Empirical Evaluation Using Social Network Data

Megumi Watanabe, *University of Nebraska-Lincoln*; Kristen M. Olson, *University of Nebraska-Lincoln*; Christina D. Falci, *University of Nebraska-Lincoln*

Survey researchers have long hypothesized that social isolation negatively affects an individual's likelihood of participating in surveys, while social integration increases the likelihood of survey participation. However, measures of social isolation usually rely on proxies that measure marginalized groups in the population or isolated groups, such as the elderly and nonwhites, not on direct measures of social isolation, that is, lack of connectedness to other persons in general or lack of connection to similar others. We use the 2008 Survey on Promoting Success among Faculty (AAPOR RR2 63.6%) to examine the relationship between social isolation and survey participation. This study examines social networks among faculty in academic departments. In this study, faculty identify the people in their department with whom they collaborate for research or consider to be friends. Importantly, nonrespondents to the study can be identified as research collaborators or friends by study respondents. Thus, social isolation measures are available on both respondents and nonrespondents. Standardized indegree (the number of connections divided by the department size) is used to measure 'general social isolation' in research exchange or friendship network. In addition, individuals' connections to other people in their department with similar group characteristics, or homophily, can be identified for a measure of 'group isolation'. Preliminary analyses indicate that standardized indegree in both research exchange and friendship network are positively associated with survey participation rates. Also, we found that gender homophily in friendship network increases the likelihood of survey participation.

Community Attachment, Social Trust and Nonresponse to a Telephone Survey

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

In a previously presented paper (Guterbock, Hubbard and Holian 2006), we showed that community attachment can be a strong predictor of individual and geographic variations in unit non-response, exceeding indicators of population density and urbanicity in its explanatory power. (Community attachment can be defined simply as the degree of connection that exists between an individual or group and its locale.) In the present research, we attempt to replicate and elaborate this finding using a larger sample and a modified measure of community attachment. The data are from a 2009 telephone survey of 2,500 adults in the National Capital Region, concerned with the experiences, attitudes, knowledge and likely behavior of the public in case of a terrorist attack. The survey used a triple-frame design, but since geographic

information is largely unavailable for the attempted cell phone numbers, the present analysis uses only landline phone numbers with known addresses. We use Census data, the survey data aggregated to the ZIP-code level, and individual demographics to identify predictors of community attachment using multi-level modeling. We then use the ZIP-code level data to identify the relationship between community attachment and various outcomes: contact, cooperation, and response rates. We find that community attachment is positively associated with response rate, operating through its separate multiplicative effects on cooperation and contact rates. We also consider a potentially important mediator of the relationship between community attachment and response rates: generalized social trust, a factor not measured in our earlier work. We discuss some theoretical and practical implications of these findings for non-response bias, and argue that community attachment and social trust should both be given more explicit attention in future research on unit non-response and non-response bias.

Survey Topic Saliency: An Examination of Potential Effects and Remedies **Johnny Blair, *Abt SRBI*; Pat D. Brick, *Westat*; J. Michael Brick, *Westat***

In this paper, we review, from a broad perspective, what is known about the potential effects of survey topic on survey quality, what approaches have been used to mitigate undesirable effects, and what directions this knowledge and experience suggest for survey design and research. The conventional wisdom that sample members' interest in the survey's topic affects their participation decisions has motivated the development of at least one theoretical model to understand topic saliency effects and their interactions with other factors (Groves, Singer and Corning 2000). Both practical experience and empirical research (e.g. Groves, Presser and Dipko, 2004) suggest that topic salience can produce differential response rates, with the potential to introduce bias when sample estimates differ by subgroup.

Researchers have investigated whether these effects may be mitigated by features of the questionnaire (e.g. Brick et al. 2012) or by survey design. For example, Schwartz et al. (2006) explored whether topic interest effects could be counterbalanced through the use of sample quotas and incentives. Beyond unit response, there is evidence that salience at the item level may affect item nonresponse (Adua and Sharp 2010) and measurement error (Stern, Smyth and Mendez 2012).

There are several features of a survey may potentially affect or mitigate topic saliency effects, including:

- Survey sponsorship
- Survey design
- Data collection mode
- Incentives
- Questionnaire design features
- Questionnaire length

There is no model that includes these multiple factors, or an empirical study that addresses the entire range of design characteristics. The synthesis of the research literature and survey methodology reports that we have undertaken will inform both future research and current practice.

Partisanship and Nonresponse in Political Polls

Leah M. Christian, *Pew Research Center*; Michael Dimock, *Pew Research Center*; Danielle Gewurz, *Pew Research Center*; Scott Keeter, *Pew Research Center*; Jocelyn Kiley, *Pew Research Center*; Alec Tyson, *Pew Research Center*

Nonresponse in social and political surveys continues to grow. Although nonresponse can often be distributed at random and is not usually a good indicator of bias in a survey, one issue of particular concern to political pollsters is whether Republicans and Democrats respond to surveys at similar rates and whether political events may differentially influence motivation to respond to a survey request. This paper will explore whether there is evidence of differential nonresponse between Republicans and Democrats in political polls. We will draw on a major study of survey nonresponse conducted in 2011, as well as data from four surveys conducted by the Pew Research Center from September through early November 2012 that included more than 10,000 respondents. Since the actual partisan affiliation of nonrespondents is unknown, this paper will use two approaches to examine the potential for political nonresponse bias. First, samples files will be matched to external databases to assess how well partisan affiliation compares among respondents and nonrespondents. In addition, a geographic analysis will use county level presidential vote data from the 2012 election to see if people in areas that voted at higher rates for Mitt Romney (“red counties”) responded at similar levels to people living in areas that voted at higher rates for Barack Obama (“blue counties”).

Tracking and Re-engaging Respondents for Follow-Up Research: A Methodological Examination of Two Research Studies

Anna Sandoval, *American Institutes for Research*; Celeste Stone, *American Institutes for Research*

Securing participation in follow-up studies is completely conditional on a study’s ability to relocate the original participants. Recent advances in technology and the availability of less expensive methods for relocating sample members have led several recent efforts to revitalize previously “decommissioned” longitudinal studies, recontact participants for program evaluations, and reconstitute studies initially designed as cross-sectional studies to allow for the investigation of complex social phenomena. However, some individuals are harder to relocate than others, and tracking biases may result if location propensity is systematically related to outcomes of interest. This paper uses data from two recent studies to examine the effectiveness of strategies for relocating and reengaging study participants after long periods of noncontact. In Study 1, researchers sought to relocate participants of a postsecondary program aimed at increasing the ethnic diversity in the aquatic sciences and who last participated in the program anywhere from 5-20 years previously. Study 2 is a pilot test assessing the feasibility of finding and reengaging a national, representative random subsample of Project Talent participants (now ages 65-70) who had not been contacted in 37 to 51 years. Both studies used commercially-available databases for tracking and prepaid incentives. This paper summarizes (1) the results of the tracking activities, including careful examination of key factors on tracking success and the effectiveness of low-cost strategies for locating participants, and (2) the utility of unconditional prepaid incentives on response rates after long periods of noncontact. This study also explores the types of individuals who are underlocated. Results from this paper will be used to inform others interested in revitalizing studies about possible biases associated with tracking and reengaging participants after a long hiatus.

Polling Around the World

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). Such tendencies are found even in countries that have traditionally been portrayed as American allies, such as Great Britain and Poland. Strengthening of 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 will be 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.

When Undecideds Decide It All: The Effect of Unreported Opinions on the Results of Pre-Election Polls

Mohamed Abouelela, *Faculty of Economics and Political Science*; Magued Osman, *The Egyptian Center for Public Opinion Research (Baseera)*

Pre-election polls were not much welcomed as a new practice in Egypt. None of the pre-election polls about the last Egyptian presidential elections in the first round (conducted on 23-24 May 2012) succeeded in predicting Mohamed Morsi (current Egyptian president) to be one of the winners to the second round of the election. Opinion polls were strongly criticized of being politically biased and unscientific; three presidential candidates filed a case to ban pre-election opinion polls in Egypt. This paper analyzes the discrepancy between the results of the pre-election opinion polls conducted by the Egyptian Center for Public Opinion Research (Baseera) before the first round of the Egyptian presidential elections and the actual results. The analysis suggests that a key factor in explaining this discrepancy is the characteristics of the undecided voters. Respondents who prefer Islamic candidates were more likely not to name their preferred candidate than other respondents do. Another factor affected the quality of the pre-election polls' predictions was neglecting the trends of voters' preferences accompanied by a long lag time between the last poll and the election date. Finally, the paper suggests a corrective procedure that incorporates the characteristics of undecided voters into the prediction of the election winner(s).

Does Data Collection Method Affect the Results of the Post-Election Polling in Egypt?

Hanan Girgis, *The Egyptian Center for Public Opinion Research (Baseera)*; Magued I. Osman, *The Egyptian Center for Public Opinion Research (Baseera)*

After the Egyptian revolution, public opinion polls became one of the important means to measure the political orientations of the citizens. For the first time in Egypt public opinion centers performed pre-election opinion polls to discover the presidency election candidates who Egyptians would vote for. Other centers performed post-election polls to analyze the characteristics of the voters of each candidate. Some of those centers used phone polls to collect data and others used face-to-face interviews. A great debate was raised about the appropriateness of using the phone polls to collect data in a country with a population that suffers from low demographic characteristics like Egypt. The new democratic experience in Egypt put public opinion surveys among the most important instruments to keep progress in democracy track. Employing public opinion polls in drawing the political map in Egypt is also new. This forces pollsters to test the different methodologies and tools they use in their data collection methods. This paper aims to discover whether, in Egypt case, data collection method affect the results of voting for different candidates and to reveal whether the effect, if any, occurs in certain population groups or it is a random effect. The paper analyzes data collected by a private independent public opinion research center that collected data on the candidates for whom the Egyptian women voted using phone polls and face-to-face survey. Both the poll and the survey were performed in the same period using national representative samples of women and both of them collected data on the main characteristics of the respondents.

Indicators of State Legitimacy in Afghanistan

Nina R. Sabarre, *D3 Systems*; Samuel Solomon, *D3 Systems*; Timothy Van Blarcom, *D3 Systems*

State legitimacy is critical for policy implementation in Afghanistan, as its absence requires the central government to devote resources to maintaining sovereignty against an increasingly bold and coordinated insurgency rather than effective governance. A state is considered “more legitimate” the more it is perceived by its citizens as rightfully holding and exercising political power (Gilley 2006). With legitimacy in hand, the Afghan central government is more likely to effectively implement policies that Afghans consciously accept. This paper contributes to the discourse of state legitimacy through a quantitative analysis of variables that influence indicators of state legitimacy. In April 2012, the Afghan Center for Socio-Economic and Opinion Research (ASCOR) fielded a survey commissioned by D3 Systems, Inc. among 2,039 individuals across all 34 provinces in Afghanistan. This survey measured public perceptions of general living conditions, performance of the central government, reconciliation with the Taliban, and recent events. Working with 125 different variables, the authors of this paper use logistic regression models to isolate variables (such as: region, security, opinion of Taliban, income, religion, socio-economic status, etc.) in order to understand their influence on state legitimacy. Although a number of variables affect how Afghans perceive the legitimacy of their government, this analysis concludes that the rating of the security situation is the most powerful factor affecting perceptions of legitimacy.

South Sudan: Evolving Opinions After a Year of Independence

Brian M. Kirchhoff, *D3 Systems*; Samantha Chiu, *D3 Systems*; Matthew Warshaw, *D3 Systems*

D3 Systems of Vienna, VA fielded surveys in South Sudan in November 2011 and December 2012. These surveys of South Sudan measure public opinion as it relates to the most important issues facing this new country. This paper analyzes survey results and compares trends after one year of independence. In 2011, a honeymoon period produced positive opinions, but a year later opinions on multiple topics have shifted. The research topics include political stability, hydrocarbon policy, delivery of services and resources to a largely rural population, the HIV/AIDS epidemic, regional drought and famine, the regional spread of terrorism and a perennially contentious relationship with Sudan. In addition to improving understanding on the aforementioned issues, the surveys also capture key demographic information and include questions that measure media penetration and usage. Due to the low penetration of phones and Internet throughout the country, the surveys were conducted via face to face interviewing. Local interviewers were recruited primarily from universities and were trained for two days prior to commencing field work. The questionnaires were prepared in English and Arabic. Interviewers were required to be fluent and literate in English and at least one other. The wave 1 sample consists of 5 key cities across South Sudan, with a representative sample of the 18+ population by city, gender and age group. The five cities are Juba, Malkal, Rumbek, Yambio and Wau. The wave 2 sample was split into urban and rural sub groups; 500 interviews were conducted in the same five cities that comprise the wave 1 sample frame and an additional 500 interviews conducted in rural locations surrounding those five cities. Respondents were selected using a multi-stage random method, from PSU selection (from a proportional to population list of sampling points), to household selection (random route) and respondent selection (Kish grid).

Strategies for Increasing Response Rates

Use of Smart Phones/Text Messaging to Increase Response Rates

Piper DuBray, *ICF International*

INTRODUCTION: Survey response rates have greatly declined in the past decade, causing researchers to seek new ways to increase participation. The Connecticut Department of Health (CT DPH) and ICF International conducted two pilot studies in 2012 using text messages to 1) increase response rates to the Behavioral Risk Factor Surveillance System (BRFSS) cell phone survey, and 2) increase participation in the BRFSS Non-Response Web Follow-up.

METHOD: To evaluate the impact of an advance text message on survey response, the CT BRFSS cell phone sample was divided into 3 groups: Group 1 was sent a text asking the respondent to complete the telephone survey when called, also offering a \$10 incentive. Group 2 received the text invitation with no incentive offer, and group 3, the control group, did not receive a text message. The second pilot consisted of sending BRFSS telephone non-responders a text message invitation to complete the survey via Web. Non-responders were divided into 2 groups: Group 1 received 2 text messages inviting them to participate in the Web survey and offered a \$10 incentive for participating. Group 2 was sent the text invitations, without an incentive.

RESULTS: Early results show that text invitations to the Web survey do not have a significant effect on response rates. Initial results of advance texts to cell phones show a 2% increase in CASRO over the control group, while advance text with incentives show a 3% increase in

CASRO over the control group. We will conduct further analyses after all data has been collected, and determine whether this increase in response rate is significant.

CONCLUSION: Based on preliminary results, text messages as a tool to increase response had mixed results. Advance text messages increased participation in a telephone survey, but text messages to BRFSS non-responders were ineffective in increasing Web survey participation.

The Use of Email, Text Messages, and Facebook to Increase Response Rates Among

Adolescents in a Longitudinal Study

Anna Fleeman, *Abt SRBI*; Kimberly Francis, *Abt Associates*; Tiffany Henderson, *Abt SRBI*; Michelle Woodford, *Abt Associates*; Marlena Jani, *Abt SRBI*

Over the course of two years, more than 1,600 students in grades 7 through 12 were recruited to take part in a three-year study assessing the effectiveness of a pregnancy prevention program. As part of the assent process, students were informed about the study and asked for name, home address, home phone, cell phone, email address, Facebook username, and permission to text. The study consisted of three 25 minute surveys: a baseline with a \$15 incentive, a 12 month follow-up (\$25), and a 24 month follow-up (\$30). The baseline survey was administered in-school either online or by paper/pencil, with both follow-up surveys conducted online. The majority of students were from low-income, minority households; therefore, six months after the baseline and first follow-up surveys, they were asked to confirm or update their contact information in an online five minute tracking survey. Initially, the first tracking survey promised a \$5 incentive; however, due to low response, it was increased to \$10 as well as text messages were sent as reminders. Additionally, phone calls were added both as a reminder and as a mode to complete the short tracking survey. For the first follow-up, invitation and reminder contacts consisted of a minimum of six emails, three text messages, three letters, and six phone calls, depending on available contact information. To increase response, we decided to send Facebook messages using the standard publicly-available personal page with privacy as the utmost concern. Presented results will include response rates and demographics by contact type and timing. Further, the operational issues related to text and Facebook messaging will be detailed. The results provide great insight as to the use of social media as well as to the retention, contact, and response rates of surveys of adolescents.

Will They Answer the Phone If They Know It's Us? Using Caller ID to Improve Response Rates

Kathy Ott, *National Agricultural Statistics Service*; Heather Ridolfo, *National Agricultural Statistics Service*; Jeff Boone, *National Agricultural Statistics Service*; Nancy Dickey, *National Agricultural Statistics Service*

Survey response rates have been declining over the last several decades. In terms of telephone surveys, this decline is often attributed to the wide availability of call screening technologies and respondents' reluctance to answer calls from unknown numbers. This has led some to posit that calling respondents from local area codes (or familiar area codes) and using identifiers that are both recognizable and trustworthy may improve survey response rates. In fact, anecdotal evidence within our own agency has suggested that this may be the case; however, research outside of our agency has produced mixed findings in regards to these claims. At the National Agricultural Statistics Service, we conducted a series of experiments to determine if the information presented on caller ID would influence response rates. Specifically, we examined

whether calling respondents using local area codes rather than out-of-state area codes and different identifiers (i.e., USDA versus Ag Counts) improved response rates. In addition, we surveyed respondents regarding their use of caller ID and its influence on their decision to answer our call. In this presentation, we will discuss the findings from this study and their implications.

Using Qualitative and Quantitative Testing to Improve Hispanic Response Rates for Online Surveys

Yelena Pens, *Arbitron*; Robin Gentry, *Arbitron*

Arbitron Inc., a provider of radio ratings data, conducted a test using a probability based address sample to recruit the Hispanic population, aged 13 and older, to complete a one week Web-based diary of their radio listening. Since hard-to-reach demos such as the Hispanic population historically have had lower participation, a qualitative study was conducted to provide insights into the Hispanic population and used to design materials used in a large quantitative study of recruitment into an online survey. In January 2012, Arbitron conducted a series of focus groups as well as face-to-face interviews with the Hispanic population in three markets. The purpose of the focus groups was to determine concerns related to the mailing materials. In particular, materials presented included mailed invitations for the Web-based diary. The face-to-face interviews were conducted in the form of a usability study in order to provide insight into the user experience of the Web-based diary. The mailing materials as well as the Web-based diary were translated in Spanish, thus the participants were able to select an English or Spanish version of the online diary. In October 2012, Arbitron conducted a pilot study of the online diary for the Hispanic population. The feedback from the qualitative study helped to design advanced notices, mailing invitations, and pre-recorded blast messages for the Web-based diary. The usability study helped to re-design the Web-based diary that was previously used for a pilot study of the general population. In this presentation, we will present the results from the qualitative and quantitative studies. In addition, we will present the optimal strategy for mail-based recruitment for an online survey of the Hispanic population.

Survey Reminder Method Experiment: An Examination of Cost Efficiency and Reminder Mode Salience in the 2012 N-MHSS Locator Survey

Matthew G. Anderson, *Mathematica Policy Research*; Barbara Rogers, *Mathematica Policy Research*; Karen CyBulski, *Mathematica Policy Research*; John Hall, *Mathematica Policy Research*; Cathie E. Alderks, *SAMHSA*; Laura Milazzo-Sayre, *SAMHSA*

Encouraging survey completion rates in a cost-efficient manner is typically a challenging endeavor. This paper will use data from the 2012 National Mental Health Services Survey (N-MHSS) Locator Survey to examine whether one type of respondent reminder is more cost-efficient than another. The 2012 N-MHSS is sponsored by the federal government's Substance Abuse and Mental Health Services Administration and includes 22,455 mental health facilities across the United States. Data for this survey were collected using the Web mode with computer-assisted telephone interviewing (CATI) follow-up over a four-month period. Using 4,300 randomly selected facilities equally divided between treatment and control groups, facilities received one of two types of reminder. A specialized reminder letter was mailed first class to the control group and to nonresponders in the non-experiment sample. The treatment group received CATI reminder calls, starting on the same day that the letters were mailed. A two-week field period was established to complete the reminder calls and to allow the letters to

arrive at facilities. Our findings will include an analysis of the percentage of facilities that completed the survey during or shortly after the reminder period and an examination of facility characteristics that might affect the completion rate, in addition to analyzing costs. The results of this experiment will help determine whether a particular reminder method is more efficient, both in cost measures and completion rate, and can help inform the survey research field of evolving trends in respondent behavior and reminder mode salience.

The Role of Blogs in Public Opinion Research Dissemination

The Survey Geek

Reg Baker, *Market Strategies, Inc.*

Reg Baker launched his blog The Survey Geek in 2005 as a way to share news and information about survey methods with his colleagues at Market Strategies International. As those colleagues shared posts with clients and others outside the company it morphed from an internal blog into a public blog. Its content also evolved from a focus on survey methods to broader commentary on the evolution of new research methods of all kinds. The blog's original intent was to educate and while some posts still have that theme it more often is commentary on how the research industry is changing, whether that's for the good or ill, and is especially disrespectful of the hype that dominates too much of the so-called "NewMR." Reg is the former president and COO of Market Strategies International where he now works as a part-time consultant.

LoveStats

Annie Pettit, *Conversion*

Annie Pettit launched her LoveStats blog four years ago after leaving a full-time job to pursue her own interests. The blog began simply as a way to stay active in the market research arena, even though she was not part of global company, but grew into much more. It became a place to clarify fuzzy thoughts, disagree vehemently with traditional opinion, pursue hot-headed rants, share insights into new methodologies, and show others that you can have a little research, statistics, and baking fun along the way. As the blog became more popular, it led to many unexpected opportunities that a behind the scenes researcher rarely gets to participate in. Annie is the Chief Research Officer of Conversion Strategies and Vice President, Research Standards at Research Now, specializing in social media market research, survey research, and data quality.

SurveyPost

Adam Sage, *RTI International*

RTI International's SurveyPost is a blog comprised of future-oriented researchers in the fields of survey methodology, health communications, and statistics and informatics. Contributions are geared toward evaluating and understanding emerging technologies and concepts as sources of social and behavioral data, and tools for data capture. Emerging from research and development initiatives in communication platforms, such as Facebook, Twitter, and smartphones, and concepts such as crowd behavior, SurveyPost is intended to communicate with and engage the research community in ways that promote the spread of innovative research on the very platforms we investigate. Recognizing the difficulty in publishing cutting-edge research that keeps pace with the rate of technological development, SurveyPost

researchers view blogs and other forms of social media as critical mechanisms for promoting timely discussion of our research to ensure that the state of science is in line with the state of technology. Adam Sage is the editor of SurveyPost and is a research methodologist at RTI International.

The Caucus

Marjorie Connelly, *The New York Times*

The New York Times website has more than 60 blogs dealing with news and politics, business and finance, technology, culture and media, health and education, and style and leisure, sports and opinion. Most are group blog sites, written by a mix of staffers and freelancers; others are blogs by individuals. Marjorie Connelly, as an editor on *The Times*' News Surveys and Election Analysis Desk, works to coordinate the multi-platform coverage of surveys. She has been contributing to The Caucus blog, which has offered news and analysis about government and politics, since February 2007. She writes items about *Times*/CBS News polls, those released from other organizations, and other survey related news. In addition, Marjorie posts items to the local City Room blog that concentrates on news about New York City and to blogs from the business section dealing with personal finance and health care. But it's not all politics and hard news: she contributes to the sports blogs dedicated to the Olympic Games, the N.F.L., college sports and major league baseball. The Caucus and other blogs are useful ways to disseminate survey results that may not merit a full story, but are be interesting or entertaining. However, inclusion on a blog does not preclude an item from appearing in the print version of the paper. *The Times*' own surveys are often teased with partial releases on The Caucus during the afternoon ahead of the full poll release in the evening. And polls released after *The Times*' print deadlines now have a place to appear.

FreeRangeResearch

Casey L. Tesfaye, *American Institute of Physics*

Blogs are particularly important in the current research environment. Excitement abounds over the terabytes of data freely available for analysis online. This has led to a rapid rise in data science and experimental analytic strategies. The survey community has been understandably slow to embrace these developments. From a perspective of Total Survey Error, in a field where a few percentage points can have far-reaching consequences, experimental methodologies seem downright irresponsible. It is important both to advocate for our abilities and continued relevance as a field and to carefully examine the strengths and weaknesses of new methods. Casey uses her blog FreeRangeResearch as a space to experiment with ways in which these research methodologies can coexist and even learn and gain from each other. As the voice behind FreeRangeResearch, Casey aims to explore the quickly evolving field of social science research in a methodologically grounded way. She tries to maintain an up-to-date listing of relevant blogs and research tools, share high quality articles from a range of disciplines, report from a range of speakers and events, and explore intersections that she comes across in her own research.

Kumarrao.net and Survey Practice

Kumar Rao, *The Nielsen Company*

Kumar started his blog www.kumarrao.net about three years ago as a window to, what he calls his "thinking world." He saw this as a venue to not only showcase his research activities and

interests, but also to network and connect with like-minded folks who share his research interests. Folks from various disciplines and countries have contacted Kumar to ask for a copy of his papers and/or share their opinion about his research. He has also ended up working with some of them. Kumar feels, when done right, blogs can serve as a gateway to particular communities of supporters, learners, and peers. The caveat here is “when is it done right?” What does it mean? Is it the ability of a blog to differentiate itself from the millions out there? What does differentiation mean in this context? How can bloggers attempt to differentiate their blogs with the thousands of spam ones that are out there? Is it in quality or quantity of the content in the blog, on top of a good advertising strategy, that can facilitate differentiation? With his recent appointment as co-editor of AAPOR’s blog-style publication *Survey Practice*, Kumar feels an even stronger sense of my social obligation to serve the larger community of survey and public opinion researchers. A recent AAPORnet post from former *Public Opinion Quarterly* editor Peter Miller describes the role of a journal editor which Kumar believes also applies to other distributed content sources such as blogs. He wrote “editors should not give themselves the license to dictate a journal’s content and should be careful stewards rather than egotists with a grand vision.” Kumar Rao is the director for the Statistical Center of Innovation at The Nielsen Company where he is responsible for developing new statistical and computation techniques for online and mobile business initiatives.

Researchscape

Jeffery Henning, *Researchscape*

Interest in survey research and polls is surging, in part because of the rise of Do-It-Yourself survey platforms. Many business people are being asked to conduct surveys, despite having no formal training in the field.

Methodological Briefs: Internet Surveys

The Impact on Web Survey Drop-Out Rates of Page Number Progress Indicators Used Throughout, Near the End, or Not at All

Jill Walston, *American Institutes for Research*; Brittany Cunningham, *American Institutes for Research*; Rebecca Medway, *American Institutes for Research*

A common feature of Web-based surveys is a progress indicator letting respondents know how far along they are in the survey. This information can be in the form of a progress bar that steadily fills up as the survey is completed or a display of the current item or page number along with the total number of items or pages. According to Conrad, Tourangeau, & Peytchev (2004), the use of progress indicators is based on the assumption that respondents will be less likely to drop out if they see they are making progress. However, there are conflicting results on progress indicators’ effect on drop-out rates (Callegaro, Yang, Villar, 2011; Conrad, Couper, Tourangeau, & Peytchev, 2004; Matzat, Snijder, & van der Hurst, 2009). We speculate that a progress indicator might be most effective at discouraging drop-outs at the end of the survey when the respondent is close to completion. To investigate this possibility, we administered a Web-survey under three randomly assigned conditions, 1) a page number progress indicator for all 12 pages of the survey (e.g., “page 1 out of 12 pages”), 2) a page number indicator appearing only for the last 3 pages of the survey), and 3) no progress indicator. Comparing drop-outs during the first 9 pages of the survey will evaluate the impact of page numbers vs. no page numbers. Comparing drop-outs during the last three pages will allow us to consider the

impact of adding the indicator near the end of the survey. The survey is being administered to a national sample of public school principals and includes questions about Common Core State Standards. Given the ambiguity that continues to surround the effect of progress indicators we anticipate that our results will add an informative perspective on the possible impact of using a hybrid approach.

Examining the Feasibility of SMS as a Contact Mode for a College Student Survey
Scott D. Crawford, *Survey Sciences Group, LLC*; Colleen A. McClain, *Survey Sciences Group, LLC*; Sara O'Brien, *Survey Sciences Group, LLC*; Toben F. Nelson, *University of Minnesota*

As respondents use mobile devices to take Web surveys at increasing rates, researchers are finding that related technologies may also be a useful tool for communicating with these respondents. Recent work surrounding text (SMS) messages as a means of communication with respondents both at the survey invitation stage (Mavletova & Couper, 2012) and as a data collection mechanism (Brenner and DeLamater, 2012; Schober et al., 2012) has suggested promise for the communication method, while at the same time raising questions about optimal use. With this literature in mind, we focused on the processes of consent, mode of invitation, and type of URL used (due to space limitations with SMS) as we invited college students at one Midwestern university to participate in a short, rapid-response survey evaluating alcohol use over the past month. We will begin by comparing those giving consent to receive SMS messages (obtained in a baseline survey) with those who did not consent to be contacted in this way. Then, we will describe the results of a randomly assigned experiment conducted among 1367 students, in which we varied both communication type (email versus SMS) and URL composition (short, commercial "tiny URL" service versus full research domain URL). We will discuss the relationship of these treatments to both data quality indicators and substantive measures, using baseline and follow-up data in our analysis. Key measures explored will include response rates, break-off rates, item missing data rates, substantive mental health and alcohol use measures, and respondents' self-reported use of technology. Further, we will address the practical challenges of incorporating short SMS messages into a data collection protocol focusing on sensitive behaviors-- including issues related to message content length, IRB approval, consent processes, and SMS technology.

The Effectiveness of Mailed Invitations for Web Surveys
Wolfgang Bandilla, *GESIS - Leibniz Institute for the Social Sciences*; Mick P. Couper, *University of Michigan*; Lars Kaczmirek, *GESIS - Leibniz Institute for the Social Sciences*

E-mail is a common invitation mode for Web surveys. However there are limitations in conducting Web surveys of the general population because lists of all Internet users and their e-mail addresses do not exist. So it is impossible to select a random sample of e-mail addresses (compared to RDD for the telephone). One solution could be to collect e-mail addresses in another mode (e.g. via CAPI or CATI interviews). But asking for e-mail addresses may raise privacy concerns among respondents. We test whether an invitation by a mailed letter could be an alternative to the common e-mail invitation in a Web survey. In this experiment participants were recruited with the aid of the German General Social Survey (ALLBUS), a face-to-face survey using computer assisted personal interviews (CAPI) in private households, conducted in 2012. Among ALLBUS respondents who reported having Internet access at home, we asked a random third for their e-mail address: 43% provided their e-mail address, while 57% declined to do so. As a control group two thirds of the Internet users were not asked for their e-mail

address. In a follow-up Web survey, to be conducted in February 2013, the three groups of Internet users (those who provided an e-mail address, those who were asked but refused to provide an e-mail address, and those not asked for an e-mail address) will be invited to a Web survey by a mailed letter. We will examine the response rates to the Web survey among the three groups, and explore potential demographic and attitudinal differences of respondents, based on ALLBUS data. Our expectation is that those who provided an e-mail address will be the most cooperative, while those who were asked but refused will be least likely to respond to the Web survey.

A Competition Among New Graphical Methods for Eliciting Probability Distributions

David Rothschild, *Microsoft Research*

We test eight graphical interfaces that capture probability distributions from non-experts. This work stands to both improve how surveys elicit expectations from experts and allows us to successfully elicit new information from individuals, which was previously too complicated to survey. Traditional methods typically elicit probability distributions by asking for the likelihood of an outcome in a given range. More modern examples include the “ball and buckets” method of asking users to fill up buckets that represent each range with 100 balls. The new methods we propose ask users for up to six data points that define polygon-shaped probability distributions. For example, participants mark the high, low, and mid-points of a range on a ruler with their values shown both graphically and numerically. With three points set, a polygon-shaped probability distribution forms above the ruler. There is no y-axis, instead the distribution is broken up into six segments with the probability mass included in each segment indicated. The user can drag any point around freely before submitting. In various, randomly-assigned conditions we test six progressively complex methods that build from a simple point-estimate to a multi-sided shape. We compare these methods on three criteria: time of completion, effort, and accuracy of elicited moments. Faster completion times allow surveyors to either reduce monetary costs or ask more questions. Reduced effort allows users to focus more on their work, which both decreases depletion effects (which can impact results) as well as the cognitive cost of completing the survey. In an increasingly online and connected world there is a potential value to guiding non-experts to create more accurate individual-level expectations that can create more efficient choices. Further, aggregating elicited probability distributions (as opposed to simple point estimates or confidence ranges) can enhance the usefulness of forecasts for many stakeholders in many situations.

Smarter Online Panels for Smartphone Users: Exploring Factors Associated with Mobile Panel Participation

Lauren A. Walton, *The Nielsen Company*; Trent D. Buskirk, *The Nielsen Company*; Thomas Wells, *The Nielsen Company*

Smartphones currently account for nearly 50% of all U.S. cell phones and Internet usage on all mobile devices is projected to surpass that of desktop computers by 2014. Smartphones APPS also continue to rise in popularity and use across mobile platforms. With both Internet and app availability on smartphone devices, researchers have multiple methods for conducting surveys via this technology. To date relatively little research has been published about the theoretical constructs associated with survey participation on such devices. Recently one study reported a theoretical model of mobile survey participation that expanded the traditional constructs

associated with online surveys to include enjoyment and engagement. Beyond this work little is known about specific factors that influence participation in mobile surveys. In this paper we will investigate practical factors associated with a respondent's choice to participate in a hypothetical online smartphone panel where surveys are completed exclusively using mobile browsers. Specifically, using an online survey administered to a nationally representative sample of 1,000 smartphone owners, we investigate what influence survey specific factors (e.g. frequency, length, content) as well as logistical factors (e.g. personal information required, data consumption limits and GPS tracking) have on panel participation. Using a split-ballot experiment, respondents were asked to answer questions presented using either a standard 7-point Likert scale or maximum difference scaling (MaxDiff). Knowing that Likert scale formats are not optimal for smartphone Internet browsers, we explore whether similar information can be gleaned from the Maxdiff and Likert scales. Specifically, we compare both the influence rankings as well as the degree of item differentiation provided by both methods in order to assess the degree to which Maxdiff questions might provide a possibly more reliable assessment of survey factor influence and make the case that this method may be more optimal for influence questions posed on mobile browsers.

Distracted Respondents

Brian F. Schaffner, *University of Massachusetts Amherst*; Stephen Ansolabehere, *Harvard University*

The Internet is becoming an increasingly common mode for conducting survey research. While academics and practitioners have paid significant attention to evaluating the extent to which online polls are able to generate representative samples, less work has been conducted evaluating how the nature of the survey interview differs online. For example, how do respondents interact with a survey questionnaire that they are free to complete at their own pace and to what extent does the self-administered nature of online surveys affect survey responses? In this paper, we investigate this question using a series of large N online surveys conducted by YouGov, America. At the end of each survey, respondents were asked whether they had engaged in a number of activities while they were taking the survey. Half of the respondents to our surveys reported at least one distraction while taking the survey; the most common distractions included watching television, having a conversation with another adult in the room, taking a break, answering email, or taking a phone call. We combine answers to this question with data on how long the respondents took to answer each question in order to determine when distractions occurred. These data allow us to examine not only when respondents become distracted, but also whether response patterns are altered after these distractions. Ultimately, the findings from this study provide an improved understanding of how to best administer and analyze data from online surveys.

Are Response Rates to a Web-Only Survey Spatially Clustered?

Lee Fiorio, *NORC at the University of Chicago*; Michael Stern, *NORC at the University of Chicago*; Ned English, *NORC at the University of Chicago*; Ipek Bilgen, *NORC at the University of Chicago*; Becki Curtis, *NORC at the University of Chicago*

Over the past decade, researchers have learned a great deal about the design and implementation of Web surveys. However, to date, we have virtually no empirical information about the role space and place has in influencing the error associated with Web-only surveys. The two types of error most often discussed when considering Web surveys are coverage and non-response; both of which are typically indicated as reasons for low response rates in these

types of surveys. One way to pursue this issue of place is to use Geographic Information Systems (GIS) to spatially-model survey response rates. This will allow us to understand the impact of location on error in Web surveys. In this paper, we attempt to examine this gap in the literature by assessing the spatial clustering of response rates to a general population Web-only survey. The data come from a random, Address- Based Sampling approach using the Delivery Sequence File (Valassis version) where respondents received a postal letter with a URL. We calculate response rates at several geographic scales, including county, state, and region, to determine the extent to which response rates are spatially clustered. While controlling for ACS demographics, Internet availability, and postal characteristics, we then build a spatial lag model to measure spatial dependence of response rates observed. Preliminary findings show clusters of low response rates in the South that cannot be accounted for by other variables in the model.

Interviewers and Interviewing

Frequentist and Bayesian Approaches for Comparing Interviewer Variance Components in Two Groups of Survey Interviewers

Brady T. West, *Institute for Social Research, University of Michigan*; Michael R. Elliott, *Institute for Social Research, University of Michigan*

Survey methodologists have long studied the effects of interviewers on the variance of survey estimates. Statistical models including random interviewer effects are often fitted in such investigations, and research interest lies in the magnitude of the interviewer variance component. One question that might arise in methodological investigations is whether or not different groups of interviewers (e.g., those with prior experience on a given survey vs. new hires) have significantly different variance components in these models, which could mean, for example, that certain groups might benefit from additional training (in hopes of minimizing the mean squared error of survey estimates). Unfortunately, popular frequentist approaches to making inferences about interviewer variance components in hierarchical generalized linear models (HGLMs) for non-normal survey variables have several limitations. These include reliance on asymptotic theory, questionable properties of classical likelihood ratio tests when pseudo-likelihood methods are used for estimation, and a failure to account for uncertainty in the estimation of features of prior distributions for model parameters. This paper compares and contrasts alternative approaches to making inferences about differences in variance components between two independent groups of survey interviewers. A Bayesian approach is proposed that circumvents many of the problems associated with alternative frequentist approaches. The Bayesian approach and alternative frequentist approaches are applied to an analysis of real survey data collected in the U.S. National Survey of Family Growth (NSFG), and results suggest that inferences can vary depending on the approach used. Examples of software code that can be used to implement both approaches in practice will be provided as a part of the presentation.

Interviewer Perceptions and Data Collection Outcomes on a National Multi-Mode Study

Micah Sjoblom, *NORC at the University of Chicago*; Vicki Wilmer, *NORC at the University of Chicago*; Marietta Bowman, *NORC at the University of Chicago*; Peter Hepburn, *NORC at the University of Chicago*

The National Survey of Early Care and Education (NSECE) employed a multi-mode design including a national in-person data collection effort. The complexity of managing multiple combinations of samples, questionnaires and respondent types created greater needs for customized combinations of paradata and cost management data to steer outreach efforts. To establish another set of objective information reflective of experiences “on the ground,” interviewer observations were collected for specific types of cases such as considering eligibility for unscreened households or gauging whether or not a respondent would complete the questionnaire based on past contact attempts. For the observations, interviewers were instructed to complete case reviews for certain types of cases at different stages of data collection and assign a code that best matches the current status of the case. The observation process included the evaluation of previous contacts, the determination of the level of difficulty perceived in achieving case resolution and the identification of barriers to cooperation. Interviewer observations and perceptions were then used in aggregate to identify patterns and develop targeted strategies for working particular types of cases. The interviewer observations captured at numerous points during the course of data collection will allow us to further examine the possibility of using such information in systematic ways to better target effort. For this presentation we will evaluate the quality of these interviewer observations by comparing coded interviewer assessments with the additional effort expended in future contact attempts as well as the final case status outcomes assigned to these cases at the end of data collection. These comparisons will be discussed in terms of how effective the initial interviewer observations were at determining final case level outcomes and the level of agreement between interviewer observations and finalized case status assigned at the end of data collection.

Factors Influencing the Quality of Interviewers’ Observations of Respondents’ Gender in Telephone Surveys

Susan K. McCulloch, *Joint Program in Survey Methodology*; Frauke Kreuter, *University of Maryland, JPSM & IAB*

According to a 2011 survey, 68% of all U.S. organizations that conduct telephone surveys collect respondent gender data by requiring interviewers to observe and record whether they are speaking with a male or female based solely on respondents’ voice. These gender observations are often made early in the survey as part of the introduction and screening process – thus, providing limited acoustic cues to inform judgments. Researchers rely on these gender observations to: (1) understand attitudes and behavior; (2) screen for study eligibility; (3) determine skip patterns; (4) contribute to nonresponse assessment and adjustments; (5) inform post-stratification weighting; and (6) design experiments. Despite this fundamental role in research, literature suggests observational data is often flawed. In fact, analysis of the quality of one firm’s interviewer gender observations found an overall misclassification rate of approximately 8% (McCulloch et al., 2010), and higher among certain groups such as women and African-Americans. Given this, can we identify some predictors of observational errors? Moreover, how can we begin to improve the quality of gender observations in telephone surveys? The goal of this paper is to identify structural features (such as length of exposure to a respondent’s voice and the buzzing sound of a centralized phone room) in addition to

interviewer characteristics as predictors of errors in interviewer observations of gender. Utilizing existing recording of survey interviews, the experimental research addresses the following questions: (1) Does allowing more time to disentangle gender cues improve observations?; (2) Does a noisy phone room contribute to errors in observations?; (3) Are there characteristics of the interviewer and/or respondent that are significant covariates of error in interviewer observations? Using the recent paradata work and linguistics literature as a foundation to design this lab experiment, the results of this paper provide information for improved collection methods of observational data.

Shocking Misbehavior by Face-to-Face Interviewers: The 2008 ANES Office Recognition Questions

Hector Santa Cruz, *Stanford University*; Jon A. Krosnick, *Stanford University*

In 2008, for the first time in the study's history, the American National Election Studies (ANES) made audio recordings of survey respondents' answers to four open-ended quiz questions assessing political knowledge. In the past, interviewers typed transcripts of the answers while respondents were speaking; however, inspection of these transcripts revealed that interviewers usually did not follow their instructions to provide literal, word for word verbatim transcriptions. ANES made audio recordings of respondents' oral answers in 2008, to see whether more exact transcriptions of respondents' actual utterances might lead to more reliable and valid coding. These recordings were invaluable in finding remarkable deviations by interviewers from their instructions, in many cases invalidating the answers provided by the respondents. Interviewers both increased and decreased the likelihood of a respondent answering correctly by giving hints, answers, comments, choices, mispronounced names, and even completely different names. The Political Psychology Research Group (PPRG) at Stanford worked with the audio transcriptions of all respondents and coded deviations to determine their frequency and effects. Frequency of deviations include how many and how often interviewers deviated. Effects include identifying if helpful deviations led to correct answers and if hurtful deviations led to incorrect answers. Without these audio recordings, we would never have been able to discover these deviations. Our findings reveal interviewer misbehavior and show how they affect the data that countless scholars use nationwide. While the costs of survey research have increased, this study shows that the additional cost of producing audio recordings justify the benefits of increased accuracy.

Audio-Recording of Verbatim Thinkalouds: A Solution to the Problems of Interviewer Transcription?

Patrick Sturgis, *University of Southampton*; Nick Allum, *University of Essex*; Rebekah Luff, *University of Southampton*

Recent attention in the survey methodological literature has turned to the quality of the coding that has conventionally been applied to verbatim response data. Verbatim responses require respondents to express their thoughts about some topic or attitudinal object 'in their own words'. This type of question has been argued to provide potentially richer data than standard closed-format response alternatives, which are not constrained by the (generally implicit) a priori framing of an issue by the researcher or question writer. However, the potential benefits of the verbatim format are often undermined by the quality of the procedures that are used to record and code them. In this study, we use a split sample design incorporated in a nationally representative face-to-face survey to assess the effect of audio-recording verbatim responses,

compared to the standard approach of requiring interviewers to type the responses into the laptop computer as they are enunciated. We compare the responses obtained from each random half of the sample on a range of different measures of data quality as well as the distributions obtained when they are coded to the same underlying frame.

Designing Effective Rating Scales

A Comparison of Branched and Unbranched Rating Scales for the Measurement of Attitudes in Surveys

Emily E. Gilbert, *University of Essex*

The choice of question response format is an important one and has wide implications for reliability and validity. One relatively recent innovation has been the use of 'branched' formats for Likert scales. In this format, one firstly asks the respondent about the direction of their attitude, and then using a follow up question, measures the intensity of the attitude (Krosnick and Berent, 1993). The potential advantage of this method is to reduce cognitive burden on the respondent, thereby permitting data of higher quality to be extracted. The potential disadvantage is in administration time. A key question is whether potential costs of adopting this method in face-to-face surveys justify any gains in reliability. This paper uses data from wave 3 of the Innovation Panel, a subsample of Understanding Society, a longitudinal panel of 40,000 British households. A split ballot experiment was embedded within the survey, allowing for a comparison of responses between branched and unbranched versions of the same questions. In particular, reliability of both versions was assessed, as well as differences in the time taken to answer the questions in each format. In a total survey costs framework, this allows us to establish if any gains in reliability are outweighed by the additional costs incurred because of extended administration times. Initial findings show evidence of response differences between branched and unbranched scales, particularly a higher rate of extreme-responding in the branched format. However, the differences in reliability between the two formats are less clear-cut. The branched questions took longer to administer than the unbranched versions, potentially increasing survey costs significantly.

Do Branched Rating Scales Have Better Test-Retest Reliability Than Un-Branched Scales? Experimental Evidence From a Three-Wave Panel Survey

Nick Allum, *University of Essex*; Emily Gilbert, *University of Essex*

The use of 'branched' formats for rating scales is becoming more widespread because of a belief that this format yields data that are more valid and reliable. Using this approach, the respondent is first asked about the direction of his or her attitude/belief and then, using a second question, about the intensity of that attitude/belief (Krosnick and Berent, 1993). The rationale for this procedure is that cognitive burden is reduced, leading to a higher probability of respondent engagement and superior quality data. Although this approach has been adopted recently by some major studies, notably the ANES, the empirical evidence for the presumed advantages is actually quite meagre. Given that using branching may involve trading off increased interview administration time for enhanced data quality, it is important that the gains are worthwhile. This paper uses data from an experiment embedded across three waves of a national f2f probability-based panel survey in the UK (the Innovation Panel from the 'Understanding Society' Survey). Each respondent was interviewed once per year between 2009 and 2011. We capitalise on this repeated measures design to fit a series of models which

compare test-retest reliability, and range of other indices, for branched and un-branched question forms, using both single items and multi-item scales. We present the results of our empirical investigation and offer some conclusions about the pros and cons of branching.

Controlling for a Response Order Effect in Ranking Items Using Latent Class Choice Modeling

Ingrid Vriens, *Tilburg University*; John Gelissen, *Tilburg University*; Guy Moors, *Tilburg University*

The ranking approach is an often used method to measure human values. It is based on Rokeach's idea that 'a value is an enduring belief that a specific mode of conduct or end-state of existence is personally preferable to an opposite or converse mode'. The benefit of this method compared to the rating approach is that it forces respondents to choose between given choice options, while in a rating task respondents can rate all choice options as equally important. A disadvantage of the ranking approach is the occurrence of a response order effect. This means that choice options have a higher probability of being chosen just because of their placing in the list instead of their actual content. This may be the consequence of satisficing behavior (meaning that instead of looking for an optimal solution, respondents tend to go for the first acceptable option they see) and is especially visible in longer lists of items, although previous research has also shown this effect for questions with only three items. Whereas in earlier studies this effect was only detected, we show an approach on how to actually control for this effect. To do this we use the latent class factor model (with an especially designed Choice module that makes it easy to appropriately analyze choice data) and include the response order effect as an attribute of the choice. We examine the changes in model parameters when the response order effect is being controlled for or not and specifically whether this changes the effects of covariates on the content factor. We illustrate our approach with data that were gathered by implementing a small experiment in the LISS panel research project, which provides a panel that is based on a representative sample of the Dutch population.

Measurement of Self-Rated Health Among U.S. Hispanic Populations

Mingnan Liu, *University of Michigan*; Sunghee Lee, *University of Michigan*

Self-rated health (SRH) is a widely used survey item for monitoring current population health and predicting its future. While its importance is evident in survey practice and substantive research, there is no clear principle for its measurement approaches. In fact, SRH is operationalized in various forms in different surveys. Yet, it appears implicitly assumed in substantive research that SRH measured in different forms provides equivalent measurement properties. In this study, we focus on the U.S. Hispanic population and compare measurement properties of SRH implemented in four different surveys: the Health and Retirement Study (HRS), the Hispanic Established Populations for the Epidemiologic Studies of the Elderly (Hispanic-EPESE), the National Health Interview Survey (NHIS) and the National Latino and Asian American Study (NLAAS). SRH in these surveys differed by response scale (5 versus 4 point scale), question order (before versus after specific health items) and question content (overall general health versus specific domain's general health). Moreover, the item was translated differently in Spanish. This study will analyze 1) the distribution of SRH, 2) the well-known relationship between SRH and specific health conditions and between SRH and health care utilization, and 3) the level of SRH utility for predicting mortality. We will compare these estimates for Hispanics across all applicable surveys by interview language. We will also use

the non-Hispanic White sample in HRS as our benchmark group in assessing the measurement properties.

Rating Scale Design in Developing Countries: A Split Ballot Experiment in Ethiopia

Charles Lau, *RTI International*; Emilia Peytcheva, *RTI International*

Due to the growth of cross-cultural surveys, questionnaires developed in the U.S. and Europe are often translated and used in developing countries in Africa, Asia, and Latin America. These surveys often include rating scales to measure attitudes. However, there is little empirical evidence about the reliability and validity of rating scales in developing countries, or research about the optimal design of these scales. This is problematic because rating scales are likely understood differently in developing countries due to their cultural and socioeconomic contexts. To address this gap in our knowledge, we conducted a split ballot experiment in a face-to-face survey of Ethiopian business owners (n = 608). The survey included 38 agree/disagree questions about the social and economic context of doing business in Ethiopia. We randomly assigned one of three rating scale types to each respondent: (1) Verbal scale (e.g., Completely Disagree, Somewhat Disagree, Neutral, Somewhat Agree, Completely Agree); (2) Numeric scale (1-5, with verbal labels at the anchors); (3) Branched or unfolding scale that first asked about direction (Agree, Disagree, Neutral) and then asked about extremity (Completely, Somewhat). In this paper, we investigate how rating scale design affects key indicators of data quality. Three findings emerge from our preliminary analysis. First, scale design has a statistically significant effect on the distribution of responses. Compared to verbal and branched scales, numeric scales produce significantly greater endorsement of the middle category, but less endorsement of “agree” responses. Second, branched scales produce the highest levels of within-individual variance, which suggests that branched scales are best at encouraging respondents to differentiate among response options. Third, in two independent tests of criterion validity, numeric scale designs had lower levels of validity compared to the verbal and branched designs—suggesting that branched and verbal scales produce substantially higher data quality compared to numeric scales.

Partisanship, Democracy and Political Behavior

What’s Wrong With Nevada?: The Persuasive Power of Partisanship

Andrew Smith, *UNH Survey Center*; Jennifer Dineen, *University of Connecticut*

2012 pre-election polls routinely showed concerns about the economy and jobs were the most important issues facing the public. This led many analysts to believe that for most voters, economic issues would be central to their vote, much as they were in 1992 and 1980, and that Barack Obama would be denied reelection like George H. W. Bush and Jimmie Carter. But this obviously did not happen, and raises the question – why wasn’t it the economy? Frank (2004) asked “what’s wrong with Kansas” and concluded, in part, that working class whites voted Republican, despite economic policies that worked against their interests. Obama won in states, such as Nevada, and Florida, that were disproportionately hit by the recession, but whose voters did not hold his administration responsible or did not factor the economy in to their vote. Previous research has shown that perceptions of the economy are heavily influenced by partisanship (see Evans and Pickup 2010; Marsh and Tilley 2009; Bartels 2002; Conover, et. al, 1987 and Pfeffley 1987) and perhaps partisan factors outweighed economics. This paper looks

to expand this line of research, but examines specific economic consequences of the recession (loss of a job, problems paying a mortgage, adult children living at home, etc.) in addition to the attitudinal measures that make up consumer confidence scales. Preliminary findings indicate that Republicans and Democrats facing similar economic consequences view their situations quite differently; Republicans believed they were worse off than four years ago while Democrats believed they were better off. The authors speculate that voters view their economic situation based on their partisanship, reducing the impact of economics issues at the voting booth.

Types of Moderates and Their Effect on Partisanship and Voting

Natalie M. Jackson, *Marist Institute for Public Opinion*

We know that many individuals in the American population consider themselves to be ideologically moderate, and these moderate partisans and moderate independents are often the swing voters in elections. This paper seeks to understand the people that we group into the moderate category by proposing a theory of three types of moderates: the know-nothings, those who completely ignore politics and have no substantive political beliefs; the cross-pressured, those who report their ideology as moderate because they are torn between conflicting views (e.g., liberal social views and conservative economic views); and the true moderates, those who do not want to choose a side and whose political beliefs are in between liberal and conservative beliefs. Which type of moderate an individual is should determine whether they are movable partisans (the cross-pressured), independents (the true moderates), or apolitical (the know-nothings). The types and their differing mechanisms of preference formation will be illustrated using data from the American National Election Study, and data from an original national survey. By classifying moderates in this way and explaining the origins of their opinions, I move the literature beyond the assumptions that moderates are uniformly uninterested in and uninformed about politics. They are, in reality, a complex group of individuals, many of whom will comprise the swing vote in elections.

Satisfaction and Democracy: A Possible Combination?

Mónica Ferrín Pereira, *Collegio Carlo Alberto, Torino*

Satisfaction with democracy is probably one of the most contested indicators in public opinion research. In fact, it is not fully clear that this indicator in fact reflects support for democracy, as it is normally assumed. Canache, Mondak and Seligson, for example, arrive at very pessimistic conclusions, and recommend avoiding its use in research on public opinion on democracy (Canache, Mondak, Seligson 2001). In spite of the debate over its suitability as an indicator, satisfaction with democracy continues to be widely used. The vast majority of surveys on public opinion have incorporated this item in their questionnaire, and there are rich longitudinal data on levels of satisfaction with democracy in most parts of the world. In light of this, it is pressing to understand what this item measures. This paper comes precisely at the core of this discussion. It is an attempt to deal with this classical indicator from a new perspective, which is very much influenced by psychological and marketing studies. As such, I propose a new reading of satisfaction, as applied to the concept of democracy. Two main questions are to be answered through this paper: Is satisfaction with democracy a summary of citizens' expectations towards democracy, and evaluations of their democratic systems (as proposed by psychological and marketing studies)? And, can satisfaction be applied to the concept of democracy?

Consistency of Reports of Party Affiliation and Voting Behaviour—Lessons From a UK Panel Study

Nick Moon, *GfK NOP Social Research*; John Burton, *ISER, University of Essex*

One of the hot topics about polling in the run-up to the 2012 U.S. election was the role of party identification and its use by pollsters in the weighting. One of the main points of debate is the extent to which party identification is a long-term fundamental belief, or whether it is subject to quite frequent change, and may even align itself with current voting intention. This paper draws on data from the British Household Panel Study, a major study that interviewed over 10,000 adults annually for 18 years. Each year people were asked whether they supported a political party, and how strongly they did so. The paper looks at the extent to which people gave the same answer each year, whether they made a single switch in allegiance over time, or whether they move back and forth between the parties more than once. This will provide solid information on the stability of this much-used variable. In the second part of the paper we look at another political variable—reported vote at the last general election. This question was asked at 14 of the 18 waves. The paper looks at the relationship between reported vote and party identification, thus shedding more light on how respondents perceive the party identification question, and also at the stability of the reported past vote question. As British general elections are typically four years apart, respondents answered the question ‘how did you vote in the general election of XXXX?’ at three or four consecutive waves, with the election getting progressively more distant in time. There is much debate about how reliable the past vote question is as a supposedly factual question, and the paper sheds valuable light on whether this is a stable variable or not.

Friday, May 17

8:00 a.m. – 9:30 p.m.

AAPOR Concurrent Session C

Improving Surveys With Paradata

Paradata and Coverage Error

Stephanie Eckman, *Institute for Employment Research*

Coverage research involves studying the quality of the frames from which samples are selected, and the impacts of errors in frames on survey data. Coverage is an understudied area in the survey methodology literature, due in large part to the difficulty of obtaining the necessary data about errors on the frame. Fortunately, paradata can in many cases provide the missing data needed to study coverage. This presentation highlights how paradata can be used to study coverage in household surveys. It discusses several types of frames, and the studies related to each type that have made use of paradata. The presentation also suggests additional coverage research that could be done with paradata.

Paradata and Nonresponse Error

Brady West, *Institute for Social Research, University of Michigan*

Nonresponse is a ubiquitous feature of almost all surveys, no matter which mode is used for data collection, whether the sample units are households or establishments, or whether the survey is mandatory or not. Confronted with this fact, survey researchers search for strategies to reduce nonresponse rates and to reduce nonresponse bias, or at least to assess the magnitude of any nonresponse bias in the resulting data. Paradata are now used to support all of these tasks, either prior to the data collection to develop best strategies based on past experiences, during data collection using paradata from the ongoing process, or post hoc for empirically examining the risk of nonresponse bias in survey estimates or for developing weights or other forms of nonresponse adjustment. Effective design strategies for reducing nonresponse bias will call for the collection of survey process data from both respondents and nonrespondents that are correlated with both key survey variables and response propensity. Survey managers can therefore work to identify features of sample units that can be collected for respondents and nonrespondents alike which may also be related to key survey variables and the probability of responding. However, previous studies have suggested that paradata may be prone to error, and paradata collection strategies that theoretically could reduce nonresponse bias may be impaired if the collected paradata are of poor quality. Results of simulation studies designed to examine the effects of varying levels of error in survey paradata on the effectiveness of post-survey nonresponse adjustments will be discussed.

Paradata and Measurement Error

Kristen Olson, *University of Nebraska - Lincoln*

Paradata for purposes of investigating and understanding measurement error include response timing, keystrokes, mouse clicks, behavior codes, vocal characteristics, and interviewer evaluations. This presentation will focus on the analysis of these types of paradata. It will highlight the specific analytic steps taken and issues to be considered when analyzing paradata for the purpose of examining measurement error. The presentation will also call the reader's attention to issues related to the measurement error in these types of paradata and offers take-home points for researchers, survey practitioners, supervisors and interviewers.

Paradata in Web Surveys

Mario Callegaro, *Google, UK*

Survey researchers and methodologists seek to have new and innovative ways of evaluating the quality of data collected from sample surveys. Paradata, or data collected for free from computerized survey instruments, have increasingly been used in survey methodological work for this purpose. In Web surveys, paradata can be collected at a variety of levels, resulting in a complex, hierarchical data structure. One challenge is that not all off-the-shelf software programs capture paradata, and thus user-generated programs have been developed to assist in recording paradata. Further complicating matters is how the data are recorded, ranging from text or sound files to ready-to-analyze variables. This presentation briefly discusses how paradata differs by mode, and gives guidance on how to turn paradata into an analytic data set.

Paradata to Study Response to Within-Survey Requests

have evolved significantly from the days where the sole means of data collection consisted of asking respondents to complete a standard Q and A-type questionnaire administered under a single mode of data collection. Although the traditional questionnaire remains the primary instrument for data collection in survey research, it is being supplemented with requests to collect additional data from respondents using less traditional methods. Such requests may include asking respondents for permission to collect physical or biological measurements (collectively referred to as “biomeasures”), access and link administrative records (e.g., Social Security, Medicare claims) to respondents’ survey information, switch from one mode of data collection to another mode, complete and mail back a leave-behind questionnaire in a face-to-face interview, among other requests. Such requests, which are usually made within the survey interview itself, have spawned new scientific opportunities that allow researchers to answer important substantive and methodological questions that would be more difficult to answer otherwise. For a series of requests (administrative data linkage consent, consent to biomeasure collection, data collection mode switch, and income item response) this presentation will give in-depth examples of how paradata have been used to study response to each type of within-survey request. Possible uses of paradata for purposes of identifying potentially reluctant respondents and implementing intervention strategies aimed at reducing within-survey nonresponse will be discussed.

Sampling and Data Quality Issues in Internet Surveys

The Performance of Different Calibration Models in Non-Probability Online Surveys: The Case of the 2012 U.S. Presidential Election

Clifford A. Young, *Ipsos Public Affairs*

The survey research world is changing. Gold standard methodologies such as the telephone survey are under increasing pressure due to declining response rates, increased cell phone-only households, and rising costs. Many have argued that one possible solution to this problem is the online survey. There is some evidence of this. Indeed, as a class, online polls performed well in this year’s U.S. presidential elections. However, online polls have their serious critics as well. One is that online surveys potentially suffer from ‘nonignorable error, and thus to be projectable to the population must employ adjustment, or calibration, models to eliminate their bias. However, to date, calibration models have often been treated as ‘black boxes’ by the survey shops that employ them. With this in mind, we ask one simple research question in this paper: which calibration method performs best when estimating voting intention (VI)? To do this, we will analyze approximately 160,000 interviews collected for the Reuters-Ipsos presidential tracking poll between January and November 2012, including primary, state and national races. The Ipsos poll is a blended online sample where multiple panel and nonpanel sample sources are combined. Our paper will focus on the performance of different calibration models, including propensity weighting, the use of demographic and attitudinal variables in post stratification weights, and weighting strategies at the sample source stage. 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. Finally, our validating benchmarks will include both final election results as well as the weekly market averages of VI taken from ‘pollster.com.’ In total, we will have 51 separate data points to analyze.

How Do Different Sampling Techniques Perform in a Web-Only Survey? Results From a Comparison of a Random Sample Email Blast to an Address-Based Sampling Approach

Ipek Bilgen, *NORC at the University of Chicago*; Michael J. Stern, *NORC at the University of Chicago*; Kirk M. Wolter, *NORC at the University of Chicago*

In the late 1990s there was much optimism that the Web-based surveys would become the replacement for RDD telephone interviews. For many reasons, Web only surveys have not taken precedence among different survey modes. For one, according to the 2010 Current Population Survey, about 72% of the American households have Internet access. Among these households, some individuals lack the skills to use it, are uncomfortable with it, or use it infrequently. Still, among certain segments of the population, Web-surveying has become a viable part of the lexicon of survey research. As a result, more research is necessary to understand ways to sample for Web-only surveys and examine the implications of different sampling strategies on survey estimates. In this paper, we compare two Internet sampling strategies for a Web-only survey to assess the data quality and cost-efficiency obtained via each sampling strategy. In the first sampling approach, email addresses were randomly selected from a vendor's email address sample frame. We sent the sampled email addresses a series of survey invitation emails which included the link to our survey. The second sampling strategy employed an Address Based Sampling (ABS) approach and sampled addresses from the USPS Delivery Sequence File. We sent the sampled addresses a series of survey invitation mailings which included the link to our survey, as well as the instructions on how to complete the survey. We compare respondent demographics and response distributions by sampling approach and ultimately compare the response distributions obtained via each sampling approach to a national-level benchmark (e.g. General Social Survey) to assess generalizability. In addition, we explore the results of these approaches in terms of response rates, the effectiveness of incentives, and the comparison of weighted response distributions.

Can We Effectively Sample From Social Media Sites? Results From Two Sampling Experiments

Michael Stern, *NORC at the University of Chicago*; Kirk Wolter, *NORC at the University of Chicago*; Ipek Bilgen, *NORC at the University of Chicago*

The exponential increase in user generated social media sites, where individuals can share information about themselves and their opinions, has raised questions about whether we can use them in a variety of survey capacities. As a result, there is a need to investigate whether researchers can effectively sample from the social media sites and, if so, what is the quality of the data produced? In this paper, we attempt to answer these questions by comparing and evaluating two different opt-in social media sampling experiments. The two sampling methods involve using advertisements as survey invitations on two separate social media sites: Facebook and YouTube. 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 of technology use and its entry point. We assess the 1) data representativeness using the General Social Survey as our national benchmark, 2) time taken to reach 1000 completes by social media site, and 3) the cost efficiency of these sampling strategies. In addition, we conduct a series of four incentive experiments to test the effectiveness of the different quantities of incentives. The incentive experiment is designed to achieve 100 completes from each of three incentive amounts: \$2, \$5 and \$10. In addition, a

larger treatment is designed to achieve 700 completes and to test the best-value incentive determined in prior experiments.

How Far Have We Come? The Lingering Digital Divide and Its Impact on the Representativeness of Internet Surveys

J.M. Dennis, *GfK Knowledge Networks*; Curtiss Cobb, *GfK Knowledge Networks*

Even while the Internet has become a popular tool for survey data collection, researchers have identified a number of potential problems involved in using a Web-based survey. One primary concern was sampling coverage error. For example, only 68% of American households had an Internet connection in the home as of 2006 (Pew 2012). Today, more than 78% of households have an Internet connection, but some subgroups of the population such as African Americans and Latinos are still known to be more likely to be offline than others. This phenomenon is often referred to as the “digital divide.” Despite the persisting existence of the “digital divide,” the use of the Internet for survey data collection has grown exponentially. Should survey researchers still be concerned about sampling coverage issues? This study uses data from GfK’s KnowledgePanel® to examine whether attitudinal and behavioral differences—those that cannot be accounted for with post-stratification weighting—between Internet households and non-Internet households have also persisted over time. KnowledgePanel provides Internet access and netbook computers to its panelists who live in a household without Internet access. As a result, all panel members are able to participate in surveys online, minimizing the potential error resulting from the exclusion of non-Internet users. Using data from 2008 and 2012, for each year, we compared weighted estimates that include non-Internet households to weighted estimates without non-Internet households. The analysis reveals that differences still exist between Internet and non-Internet households for a series of attitudes and behaviors that cannot be corrected for using post-stratification weighting.

Respondent Validation Phase II

Dinaz Kachhi-Jiwani, *United Sample (uSamp)*; Lisa Wilding-Brown, *United Sample (uSamp)*

In the recent years, online research has gained acceptance but the question about data quality continues to surface as technological sophistication helps fraudsters to easily bypass quality checks. Prior research by Courtright and Miller (2011) highlighted the unwillingness of respondents to share Personal identifiable Information (PII) and demographic bias as major barriers to performing validation. To that end, the research was administered in 2012 to identify any change from the previous landscape and also evaluate different techniques introduced by vendors to overcome traditional challenges. We discovered that although the new techniques have an impact on number of respondents who get validated, it also influences data quality. The demographic of respondents who were not validated were consistent with 2011 and were more likely to be those without a bank account or credit card and less likely to own their own homes. When we further associated data quality with validation status, we found that respondents who failed to validate were twice as likely to fail at least one quality check (i.e. straight-line in a grid, speed through the questionnaire or answer inconsistently). Also, the validation methodologies and process of conducting validation differs across vendors. Therefore, from a project management standpoint, it becomes imperative to account for these factors to make sure that appropriate techniques are adopted and followed by researchers. Key takeaways:

- Demographic and psychographic make-up of validated and un-validated respondents.
- Validation and its impact on data quality.
- What validation means for market researchers?
- Project management implications.

Lessons in Leadership: AAPOR Women Leaders Share Their Insights

Mollyann Brodie, *The Henry J. Kaiser Family Foundation*; Courtney Kennedy, *Abt SRBI*; Nancy Mathiowetz, *University of Wisconsin-Milwaukee*; Eileen O'Brien, *Energy Information Administration, U.S. Department of Energy*

Across research sectors, there are unique challenges and concerns for women in leadership roles. Building on the experiences shared in last year's successful professional development panel, *Considering Changing Sectors in the Research Industry?*, this session will continue the conversation, focusing on women's leadership in the research industry. This panel, organized by AAPOR's Education Committee and moderated by Angie Gels of The Nielsen Company, brings together a group of AAPOR women leaders. Sharing their real-life experiences, panelists will discuss their successes and challenges as women in research and help identify opportunities to improve personal leadership skills and effectiveness. Panelists will also reflect on changing roles and experiences of women in the research industry. The panel session will include brief commentary by each panelist and a moderated Q&A session (audience questions/comments highly encouraged). The session may be of interest to women (and men!) at all levels of leadership, from informal to manager to CEO. Expect a lively discussion reflecting the diversity of our membership and their experiences. A number of experienced and willing panelists have been identified and three to four will be invited to participate in the panel.

From Concepts to Questions

Preparing to Measure Health Coverage in Surveys Post-Reform: Lessons From Massachusetts

Joanne Pascale, *U.S. Census Bureau*; Jonathan Rodean, *U.S. Census Bureau*; Jennifer Leeman, *U.S. Census Bureau*; Carol Cosenza, *Center for Survey Research, University of Massachusetts Boston*; Alisu Schoua-Glusberg, *Research Support Services*

The Affordable Care Act (ACA) is expected to be fully implemented in January 2014 and usher in a series of reforms of the U.S. health care system. One of the most significant components of the ACA is the "Health Insurance Exchange"—a state-level marketplace of health insurance options for individuals and small businesses. While these Exchanges are still in development and states have broad flexibility in designing the programs, it is essential for the federal government to have a viable methodology in place for measuring health coverage post-reform. One opportunity for research and development of such a methodology rests in the state of Massachusetts, which in 2006 passed legislation that includes most of the features of the ACA, including Exchanges. The Census Bureau teamed with Research Support Services and the University of Massachusetts to conduct research with Massachusetts residents to explore the many pathways of enrolling in an Exchange, the language and terminology residents used when describing their coverage, and ultimately to develop standardized questions for capturing

Exchange participation and subsidization. The project was conducted in three phases: expert consultation with key individuals with years of experience in measuring health coverage at the state and federal level (focusing on Massachusetts); focus groups with subgroups for whom the Exchange was targeted; and cognitive interviews with those same subgroups. Individuals with coverage through more conventional sources were included in the testing as a control to flag possible “false positives”—reporting coverage through an Exchange that was actually through another source. Questions on the Exchange were developed and tested within the context of both the Current Population Survey and the American Community Survey, thereby providing some baseline findings for other federal and state surveys that utilize a similar questionnaire structure as either of these two surveys.

Identifying the Dimensions of Question Sensitivity: A Multidimensional Scaling Study

Christopher Antoun, *Institute for Social Research, University of Michigan*

“Sensitive questions” are questions that are likely to be seen as threatening or embarrassing. They have become more common in national surveys as researchers attempt to monitor sexual behavior or the use of illicit drugs. Despite the large body of survey research about sensitive questions, it is still unclear what makes a question sensitive because no standard definition of “sensitivity” exists. Tourangeau, Rips, and Rasinski (2000) identify three distinct meanings of sensitivity from the survey literature: intrusiveness, threat to disclosure, and social desirability concerns. To date, no one has attempted to empirically verify these dimensions. Multidimensional scaling (MDS) can help by locating stimuli, which are sensitive questions in our experiment, on a spatial configuration or “dimensional space.” The ordering of the points along dimensions allow for interpretations about the nature of each dimension. An advantage of MDS is that the research participants, not researchers, determine the number and kinds of dimensions present. We conducted an experiment to empirically identify dimensions of sensitive questions. Approximately 250 participants provided pairwise similarity judgments for 12 types of sensitive survey questions. Applying MDS to these data yielded three dimensions representing how participants thought about question sensitivity. The dimensional structure did not match Tourangeau and colleagues’ formulation precisely. Intrusiveness was the most salient dimension with questions about taboo topics such as sex at one extreme of the dimensional space and a question about exercise at the other extreme. Threat of disclosure was the second most salient dimension with questions about illicit drug use at one extreme of the dimensional space and a question about racial attitudes at the other extreme. A third dimension improved the model fit but was not related to social desirability concerns. The results indicate that there are independent and separable dimensions of question sensitivity that should be further explored.

Finding the Needle: The Challenges of Recruiting Participants for Cognitive Testing by Coverage Type in an Exchange State

Katherine R. Kenward, *Research Support Services, Inc.*; Joanne Pascale, *U.S. Census Bureau*; Alisu Schoua-Glusberg, *Research Support Services, Inc.*; Carol Cosenza, *Center for Survey Research, University of Massachusetts Boston*

When recruiting for particular respondent types, there are often challenges in finding the right individuals. Researchers can advertise for a specific characteristic but this is challenging when the trait is rare in the population or is similar to another unwanted and more common trait. Screening presents challenges and sometimes yields false negatives or positives and/or primes

respondents for specific traits. In March 2010, the Affordable Care Act was passed and in 2014 Health Insurance Exchanges will be operating in all or most states. In 2006 Massachusetts passed legislation similar to the ACA and in order to develop standardized questions on Exchange participation prior to 2014 the U.S. Census Bureau, in fall of 2011, undertook to cognitively test question sets in English and Spanish exploring terms and concepts that refer to coverage through the Exchange in Massachusetts. Finding residents with coverage through the Exchange became a complex recruitment task because only a small portion of the population is covered through the Exchange and many participants do not know the specific type or source of coverage they have. In addition, the agency that administers the Exchange (and has records on enrollees) had never before allowed researchers access to their covered population. This paper explores the challenging process of identifying respondents who cannot accurately answer screening questions about their coverage source, successfully gaining the cooperation of an agency that has records on the population of interest, the limitations and benefits of using an agency as a source for outreach to the population of interest and the limitations of and creative resources used for recruiting when faced with small populations unaware of their coverage type. In addition the paper explores the time and effort involved and the benefits of each approach.

The Establishment Survey Response Process and Measurement Error: How and Why Are They Connected?

Polly Phipps, *U.S. Bureau of Labor Statistics*; Danna L. Moore, *Social and Economic Sciences Research Center, Washington State*

The BLS Survey of Occupational Injuries and Illnesses (SOII) provides a unique opportunity to study the establishment cognitive response process and measurement error. Recent studies have cited discrepancies between SOII and State Workers' Compensation (WC) administrative claims records to support the assertion that SOII undercounts workforce injuries and illnesses. To explore reasons for discrepancies, we conducted over 50 qualitative interviews with SOII respondents from establishments of varying sizes, industries, and magnitude of differences between SOII and WC data. Our in-person interviews focused on possible errors in comprehension, retrieval, judgment, and communication associated with the respondent, records system, and business environment. We address numerous questions, including across businesses and respondents, what response processes contribute the differences? Results suggest that understanding of reporting rules and survey timing play a role in discrepancies. Our research also suggests that the business environment influences the response process.

An Overarching Process for Enhancing the Validity of Survey Scales

Hunter Gehlbach, *Harvard Graduate School of Education*

For years researchers across many disciplines have undertaken the formidable challenge of designing survey scales to assess attitudes, opinions, and behaviors. Correspondingly, scholars have written much to guide researchers in this undertaking. Yet, much of their guidance focuses on discrete steps that survey designers might take – especially statistical procedures to be conducted after pilot data are collected. This paper synthesizes several of these steps into an overarching process to facilitate the construction of questionnaire scales. Unlike previous processes, this one front loads input from other academics and potential respondents in the item-development and revision phase with the goal of achieving credibility across both populations. Specifically, the article describes how (a) a literature review and (b) focus group–interview data can be (c) synthesized into a comprehensive list to facilitate (d) the development

of items. Next, survey designers can subject the items to (e) an expert review and (f) cognitive pretesting before executing a pilot test.

The Role of Literature and Parent Voices in Developing the Child Behaviors Scale **Lauren Capotosto, *Harvard Graduate School of Education***

As a first step in developing the Child Behaviors scale, we reviewed the child learning-related behaviors literature in order to define the construct and identify existing measures that could inform our own questionnaire. Many of the measures specific to children's learning-related behaviors require either teacher (e.g., Learning Behaviors Scale; McDermott, Green, Francis, & Stott, 1999) or student respondents (e.g., Patterns of Adaptive Learning Scales; Midgley, Maehr, Hruda, Anderman, Anderman, Freeman, et al., 2000). They similarly include items that reflect positive behaviors that support and negative behaviors that hinder school success. Second, we conducted focus groups with eight parents who represented a socioeconomically and racially diverse group in order to examine the extent to which the conceptualization of child learning-related behaviors in the literature aligned with the way parents conceive of it. We used a semi-structured interview protocol consisting of five open-ended questions to ascertain how parents thought about what students broadly, and their children in particular, can do to help or hinder their school success. Third, we synthesized the literature review with focus group data. Specifically, we developed a two-column list to compare indicators that emerged from the literature and focus groups. While there were several commonalities between the ways in which researchers and parents conceptualized child learning-related behaviors (e.g., both discussed procrastination as a negative behaviors and following directions as a positive behavior), there, too, were noteworthy differences. For example, whereas the literature refers to willingness to ask for help as a positive child attribute, several parents mentioned asking for help as a negative behavior. When probed further, we learned that parents wanted their children to ask for help only after making an earnest effort to work independently through a challenge. Such distinctions informed the crafting of items in step 4.

Item Development and Expert Reviews for the Child Behaviors Scale **Sofia Bahena, *Harvard Graduate School of Education***

Item development and expert reviews were two key steps in the process of developing scales for the Family-School Relationships survey. First, we used our synthesis of the literature review and focus groups to develop items for the Child Behaviors scale. The goal was to develop items that represent indicators integral to the construct, while using vocabulary that is relevant to potential respondents (parents of school aged children). We relied on known best practices in survey design to avoid wording bias, ameliorate acquiescence bias, and ensure our items would pertain to a wide range of parents. We aimed to improve reliability by avoiding reverse scored items (Benson & Hocevar, 1985; Cacioppo & Berntson, 1994; Swain, et al., 2008) and labeling answer choices with construct-specific anchors (Tourangeau, Rips, & Rasinski, 2000), and avoiding agree/disagree statements (Fowler, 2009; Krosnick, 1999). We also tried to capture the breath and depth of what children do that is conducive (or hurtful) to their school success in order to provide face validity to our items. Once developed, the items received several rounds of feedback from the research team, with a particular eye towards clarity and comprehensiveness. To address the valence of the items, we separated our scale into two sets of questions: one for positive behaviors and another for negative behaviors. Next, we reached out to field experts and asked them to review our items; 9 out of 26 responded. These experts included researchers and K-12 school leaders. We asked them to rate our items based on clarity, comprehensiveness,

and appropriateness for a broad range of parent groups (e.g. student age, parent education level, parent whose 2nd language is English). Based on their responses, one item about homework was eliminated because it did not apply to children in the earliest grades, and several others were modified.

The Role of Cognitive Interviewing and Pilot Testing in the Development of the Child Behaviors Scale

Beth

Once item development was complete, we subjected our items to a cognitive interviewing procedure to ensure potential respondents understood the items as we intended (Willis, 2005). We conducted 40-60 minute one-on-one interviews with parents regarding the Child Behaviors items (N=5). We first asked parents to restate each question in their own words, using none of the words from the item itself, and then to “think aloud” as they came to their own answer to the question. In this presentation, I will describe how these interviews a) provided evidence parents understood our items, b) led us revise a small number of items, and c) led us to eliminate a small number of items from the scale. Given this scale was designed to be deployed in the context of a larger survey, there was a premium on keeping it as parsimonious as possible. After making final wording changes based on cognitive pre-testing, we used the SurveyMonkey website to administer our survey to two separate samples of parents (N=384; N=266) from SurveyMonkey’s unique national panel. We analyzed the resulting data using confirmatory factor analysis (CFA) to provide evidence of adequate variability, reliability and the factor structure of the scale. With the first sample, our child negative behaviors items had strong internal consistency ($\alpha = .79$), as did our child positive behaviors items ($\alpha = .82$). The fit was adequate (Kline, 2011) for a two-factor measurement model with separate latent variables capturing positive and negative behaviors ($\chi^2 = 40.82$, $p = .03$; CFI = .99; RMSEA = .04; RMSEA 90% CI = .01, .06). Finally, we replicated these results with the second large sample of parents. These findings provide confidence that the Child Behaviors scale is a valuable tool for practitioners and researchers alike.

Benchmarking Parent Perceptions of Child Behaviors with SurveyMonkey

Philip Garland, *SurveyMonkey*

At least since No Child Left Behind and continuing with Race to the Top, high stakes education standards are gaining traction nationally. And while assessors obviously find value in comparing standardized test scores across schools, the process would benefit from the ability to compare schools on a broader array of factors. To that end, surveying is one avenue to collect data about schools. While test questions can be administered in a uniform way quite easily, standardized survey questions have been elusive to date. It follows then, that the ability to compare survey data across schools would greatly benefit schools and researchers alike. Schools can ascertain with much greater certainty what their strengths and weaknesses are if they are able to compare their own data to a series of comparable schools. Benchmarking data will also allow researchers to examine between-school variation. To the extent survey administration reflects the uniformity and regularity of testing, schools ought to be able to understand and explain which areas are comparative strengths and weaknesses and scholars ought to be able to understand which areas offer the greatest explanatory power. In turn, this clarity should allow school leaders to craft plans for improvement with more confidence and allow researchers to draw broad lessons about school improvement. Of course, such an offering requires critical

mass of usage. Fortunately SurveyMonkey facilitates roughly 70 million parent interviews each year from more than 80,000 schools. With this scale at hand, schools will be able to compare themselves at quite granular levels within very small geographies. Moreover, SurveyMonkey is working with Great Schools to make information available to parents as they select schools for their children. This presentation will describe this benchmarking project using the Child Behaviors scale example and will outline the implications for both researchers and school practitioners.

The 2012 Election: Horserace Polls, Exit Polls and Poll Aggregation

Voter Mobilization Effects of Localized Pre-Election Horserace Polling Information

David L. Vannette, *Stanford University*; Sean J. Westwood, *Stanford University*

Candidate performance in pre-election polls dominates media coverage of elections and is known as the 'horse race.' Yet, we have an incomplete understanding of how coverage of the horserace may influence voter turnout. Prior research has demonstrated that there are relationships between polling numbers and political behavior, but these results are largely correlational and there are large gaps in our understanding of the boundaries of these effects. We approach the question experimentally. The most commonly cited poll effects are 'bandwagon' and 'underdog' effects; simply being made aware of the preferences of other people seems to influence some voters to support the candidate or issue that is currently leading or trailing in the polls. In this paper, we experimentally examine the influence of poll reports about the state of the horserace in a potential voter's specific congressional district on the decision to turn out to vote. Participants were sent three bogus poll reports about the horserace via e-mail in the two-week period leading up to the 2012 presidential election. Participants were randomly assigned to treatments where 1) Obama was leading in local polls, 2) Romney was leading in local polls, and 3) where both Romney and Obama were "tied." An additional control group completed the pre and post-election surveys but did not receive any e-mail messages. Nearly one thousand subjects were recruited into the experiment using a convenience Internet sample. We demonstrate that reports of a Romney advantage, or a tie, increased voter turnout for Democrats. We provide detailed analysis of the moderating effects of election salience, partisan strength and media consumption. We plan to validate voter self reports with voter file data from Catalist. Our results provide evidence on the mobilizing effects of local polls on potential voters and have broad implications for polling research and political communication.

Using Non-Probability Online Surveys for Exit Polling: The Case of the 2012 U.S. Presidential Elections

John P. Vidmar, *Ipsos USPA*; Darrell Bricker, *Ipsos USPA*; Cliff Young, *Ipsos USPA*; Julia Clark, *Ipsos USPA*; Alan Roshwalb, *Ipsos USPA*; Neale El Dash, *Ipsos USPA*

The exit poll is a staple of the modern American democratic experience. Exit polls have multiple purposes including an independent check to validate election results, a mechanism to provide content to media companies during election night, and finally copious voting data for practitioners and academics alike. Traditionally, the U.S. exit poll conducted by VNS (Voter News Service) in randomly selected polling stations is increasingly feeling the strain. First, costs are a real concern. Indeed, VNS cut nineteen states from its electoral coverage in 2012 presumably due to cost. Second, early voting in 2012 reached 40% making in loco exit polling both irrelevant and problematic in the long-term. One possible solution for these challenges

would be to conduct an online exit (or election day) poll. Online methodologies, though, have their serious detractors who cite both their non-probabilistic nature and their coverage and self-selection bias. Are such criticisms warranted? To answer this question, we will analyze data from an online exit poll conducted by Ipsos for Thomson Reuters. In total, 42,000 interviews were conducted nationally. Specifically, we will compare results from the Ipsos exit poll to the traditional VNS exit poll published by the primary news shops. As a measure of comparison, we will look at the average absolute difference (AAD) to gauge relative performance of the online exit poll with more traditional methods. Finally, our paper will also detail operationally and methodologically the IPSOS online exit poll.

Information Disconnect: Data Aggregators and Media Reporting in the 2012 Presidential Election

Fred Solop, *Northern Arizona University*; Nancy Wonders, *Northern Arizona University*

Election 2012 was a \$6 billion event in the United States. Media headlines screamed news of the closeness of the race and the rhetoric of the presidential contest suggested big changes were coming. The candidates “sparred” at the debates. Reportage about the election featured a competitive “contest” between President Obama and Governor Romney. Throughout the campaign, newspapers followed every new poll, particularly those in “battleground states,” and reported on small differences in the numbers, even if they fell within the margin of error. Voters were primed to expect a tight race in a polarized election environment. Either candidate could win. In contrast to the message promoted by media outlets, some data aggregators like Nate Silver of the 538 blog and Mark Blumenthal of *Huffington Post*’s Pollster.com aggregating a wide range of election surveys and telling a different story. Their story was one of stability rather than rapid change. Obama was favored to win throughout the election season. Nate Silver, in fact, predicted that Barack Obama had a 92 percent chance of winning the presidential contest on the eve of the election. Despite contributions by these data aggregators, papers such as *The New York Times* still spoke of an expectedly close race and seemed surprised when America awoke to the news that Barack Obama would be in office for the next four years. This paper explores the disconnect between what the “data aggregators” were saying and what the media was telling voters to expect in the 2012 election. We content code and compare messages coming from both sources of information. After characterizing the extent of the disconnect, we develop a range of theories to explain this phenomenon. Finally, we offer solutions for how media can do a better job integrating the science of polling into coverage of future elections for president.

Using Model-Based Poll Averaging to Evaluate the 2012 Polls and Pollsters

Mark Blumenthal, *Huffington Post*; Simon Jackman, *Stanford University*

Much previous research on pre-election poll accuracy, from Mosteller et al. (1949) to Crespi (1988) to Traugott et al. (2005) has focused on various scores that compare the deviation between the election results and vote preference as measured on polls conducted at or near the end of the campaign. The traditional measures of poll accuracy capture both validity (or the absence of bias) and reliability (or the absence of variance). When scores are calculated from a single poll or a small number of polls, it can be hard distinguish between the two sources of error. Also, by focusing on the performance of the final pre-election poll, these scores can create unhelpful incentives, which some believe lead to a ‘herding’ of poll results near election day (Clinton and Rogers, 2012). Model based poll averaging offers another check on polling

validity, if not reliability, through estimates of pollster 'house effects,' the tendency of some survey houses to produce estimates that are systematically higher or lower for one candidate than other houses. Once the model is corrected to the final election outcome, these house effect estimates allow for tests of several hypotheses: Were the polls of 2012 collectively biased towards one candidate or the other over the final two months of the campaign and not just during its final week? Did particular sampling methodologies or surveys modes or exhibit consistent bias? And were particular survey houses better or worse in 2012?

Model-Based Poll Averaging Over the 2012 U.S. Presidential Election Campaign **Simon Jackman, *Stanford University***

During the 2012 U.S. presidential election campaign I developed a poll-averaging model that produced daily estimates of voting intentions national and state levels, published on Pollster/HuffingtonPost.com. Using over 1200 published polls, my model correctly predicted the election outcome in every state and Obama's 332 vote Electoral College tally. I elaborate the various elements of the model: (1) reliance on historical election returns; (2) corrections for house effects; (3) correlations among states and national levels; (4) a dynamic model for day-to-day changes in voting intentions over the campaign. I report estimates of key parameters of the model (e.g., house effects, the day-to-day rate of change parameter), details as to the forecasting performance of the model and sensitivity to various model assumptions. Collectively, the polling industry underestimated Obama's two-party vote share by about half a percentage point; I examine the sources of this systematic, collective bias in 2012 election polling. Since the model produces estimates of trajectories of voting intentions in every state, I also assess the extent to which 'set pieces' of the campaign (the end of the Republican nominating process, the nominating conventions, the debates) and exogenous events (e.g., Hurricane Sandy) appears to have moved voting intentions, and variation across states in the magnitude of responses to these events.

Methodological Briefs: Cell Phones

Alternative Sample Selection and Data Collection Strategies for Balancing Cell Phone Response Distribution Across County/Region Level Geographies in a Dual Frame (Landline/Cell) Telephone Survey

Howard Speizer, *RTI, International*; Marcus Berzofsky, *RTI International*; Jamie Ridenhour, *RTI International*; Tom Duffy, *RTI International*; Tim Sahr, *Ohio State University*

The challenge of targeting smaller geographic regions in a dual frame (landline/cell) telephone study increases with the proportion of sample members that are expected to respond by cell phone. For the Ohio Medicaid Assessment Study (OMAS), completed in October 2012, 25% of the 22,000 respondents in this state-wide survey completed their interview by cell phone. Targets for the number of cell phone respondents by county were set by population totals and a probability-proportionate-to-size cell phone sample was fielded. Although this design worked well in most regions, some significant under- and over-representation occurred. In this paper, we examine both sample selection and field data collection protocols to explain these variances. We then examine various cell phone data augmentation options and sampling strategies, by modeling performance on the OMAS, to improve geographic targeting without sacrificing quality

objectives. We compare the results of these alternatives and suggest an improved design for the cell phone portion of a dual frame telephone study for achieving small area targets.

Sampling Cell Phones by Rate Center: Efficacy, Coverage and Incidence
David Dutwin, *Social Science Research Solutions*; David Malarek, *MSG*

Survey researchers who conduct telephone studies of geographies any smaller than a state have limited options available for sampling cell phone telephone numbers. Furthermore, there is presently no way to estimate the incidence one will attain by any of the methods presently available for localized sampling, nor are there any techniques available to estimate coverage of the selected target population. This paper first details the options available to researchers of local telephone studies with regard to cell phone sample selection, and explains why selection at the level of rate center is superior to other methods. Using a unique dataset that combines thousands of respondent surveys across the United States with data from the 2010 U.S. Census, aggregated to the level of rate center, we show the efficacy and potential bias of utilizing rate center for local sample selection. Finally, we offer a model by which researchers who utilize rate center can estimate the survey incidence they will attain and as well as the coverage of their target population.

To Call or Mail: Impact of Mailing Surveys Directly to Cell-Phone-Only Households in an Address-Based Frame
Vrinda Nair, *Arbitron Inc.*; Robin Gentry, *Arbitron Inc.*

Arbitron currently uses a mailed screener questionnaire sent to an Address Based sample (ABS) to recruit the non-landline portion of the population. If a respondent reports being cell phone only or cell phone mainly the household is added to a cell-phone frame and used to supplement a 2+ list assisted RDD sample. This cell-phone sample is called using our CATI system. To investigate better methods and more cost-effective ways of increasing cell phone response rates, Arbitron conducted a direct mail study in fall 2012. In our current methodology, cell phone households who supply a contact phone number at the screener stage are contacted and asked to participate in a seven-day diary survey of their radio listening. With the direct mail study, this stage was omitted and radio surveys were mailed directly to the cell phone households without first gaining consent. How does not offering a chance to refuse survey participation impact response rates? We will present the return rate results, cost-benefit analysis, as well as the analysis of the demographics of those that returned the survey to determine who we brought in with the direct mail as compared to the traditional “call and then mail” approach.

Understanding Bias in Appended Wireless Billing ZIP Code Data
Tara Merry, *Abt SRBI*; Andy Weiss, *Abt SRBI*; Mikelyn Meyers, *Abt SRBI*; Paul Schroeder, *Abt SRBI*; Kristie Johnson, *NHTSA*

Designing cell phone samples is particularly challenging for small area surveys given the lack of precise geographic information available. Rate center is currently the best information available that can be used to target a specific area. While rate center information works fairly well to target larger geographic areas, it is much less precise for targeting smaller areas such as individual counties that are served by several rate centers. New post-production processes are

available that append billing ZIP code data to cellular samples. While this information cannot be used to draw targeted sample, it can be used to stratify the sample prior to fielding. The ability to append billing ZIP code data to cellular samples has the potential to dramatically improve geographic targeting precision, thereby increasing efficiency and reducing costs. The robustness and accuracy of the billing ZIP code data should be evaluated to determine how it can be used to refine cellular samples without introducing bias or increasing coverage error. We compare respondent-provided ZIP code data with billing ZIP code data in a population survey of New York City with similar data from a national survey. Approximately one third (34%) of sampled cell records in New York City have matched billing ZIP code data compared to 42% in the national sample. We review variations in the rate at which ZIP code data are matched, the accuracy of the ZIP code information, and examine which characteristics differ between matched and unmatched cases across these studies. Results are discussed in the context of how these data could be used to develop stratified cell samples depending on the geographic area being targeted, the population of interest, and survey topic.

Cell-Phone Sampling Frames: Effectiveness and Dependability of Recent-Usage Data

Robert DeHaan, *Arbitron Inc.*

Arbitron currently uses a mailed screener questionnaire to an Address Based Sample (ABS) to recruit the non-landline portion of the sample frame. The questionnaire is used to identify cell-phone only or cell-phone mainly households to be included in a supplemental cell-phone frame. Recently, innovations have been made that allow for better targeting of cell-phone only and cell-phone mainly households through the use of data available via various sample vendors. Of interest to Arbitron in these data are activity indicators, geography, and other auxiliary information to be used in stratification. To investigate the usability of these newly available data, the test will be looking into advantages and disadvantages that will come from switching over from our current address based methodology. We aim to answer: 1) How accurate are the activity indicators and are they beneficial in increasing response rates and reducing costs? 2) Can the activity indicators be used in conjunction with respondent reported cell-phone vs. landline usage to determine differential sampling rates in a dual RDD frame approach? 3) What affect, if any, will the change in methodology (elimination of screener questionnaire) have on proportionality in historically under-represented demographics? We will present results describing the effectiveness of using usage indicators in sample selection, cost considerations and comparisons, along with an analysis of demographic proportionality in agreement while using the cell phone frame.

Recent Methodological Updates Adopted for the National Immunization Survey (NIS)

Vicki Pineau, *NORC at the University of Chicago*; Robert Montgomery, *NORC at the University of Chicago*; Bess Welch, *NORC at the University of Chicago*; Kirk Wolter, *NORC at the University of Chicago*; Stacie Greby, *Centers for Disease Control and Prevention*

The National Immunization Survey (NIS), conducted annually since 1984, has been the nation's flagship survey for monitoring vaccination coverage among 19-35 month-old children. The NIS was designed to collect data using a list-assisted RDD survey methodology of households with landline telephones and a follow-up mail survey of the age-eligible child's vaccination providers

to collect vaccination histories. Like many other surveys, the NIS is affected by declining response rates and increasing cell phone only use resulting in high survey costs and serious concerns about non-representative data. The NIS research strategy in recent years was focused on assessing the consequences of noncoverage of households with cellular phone service only, addressing declining response rates in the household interview, and maximizing survey efficiency. In this paper, we summarize the major changes implemented in the NIS design for 2011-2012 and the research results that supported adoption of these changes with the relevant survey results since the changes were made. Expansion of the NIS in 2011 to a dual RDD landline and cell frame design and implementation of weighting methods to minimize the MSE of key survey estimates resulted in few changes in official vaccination coverage estimates. The change in the NIS age definition increased the number of eligible households and decreased the required number of calls in the household survey, decreasing survey costs with no substantive effect on vaccination coverage rates. Beginning 2012, the household survey questionnaire length was reduced by eliminating non-critical parental vaccination recall content resulting in higher completion rates, decreasing survey costs and supporting expansion of the cell phone sample frame. Also for 2012, an optimum dual-frame sample was fielded to minimize the survey costs subject to geographic reliability constraints. The NIS will evaluate results using data from the National Health Interview Survey Provider Record Check completed in the same years.

Cross-Platform Measurement: User Experience With a Smartphone and Web Self-Reported Data Collection Application

Ana P. Petras, *The Nielsen Company*; Shu Duan, *The Nielsen Company*; Oana Dan, *The Nielsen Company*

The proliferation and adoption of multiple Internet-access platforms in the U.S., especially among younger and ethnic populations, has increased the need for research organizations to provide alternative methods of data collection, such as mobile applications and Web. In 2012, Nielsen conducted a study in two local demographically diverse market areas in the U.S. to assess the viability of using a cross-platform mobile and online application (Whatcha Watchin'?) to capture television viewing. To maximize coverage of smartphone and non-smartphone users, iOS (iPhone/iPad), Android and online (website) versions of the application were developed and made available to participants. Here we present the end-to-end study experience through the eyes of the user and focus on recruitment, app download and general use of this data collection tool. The user experience was gathered through a follow up online survey and a set of one-on-one interviews conducted shortly after the end of data collection period. Survey respondents included those who accepted, registered into the App/Web and submitted at least one viewing entry; those who accepted and registered but did not submit any viewing entry; and respondents that accepted over the phone but didn't register into the App/Web. Key findings on the effectiveness of the recruitment materials, reasons for participation, interaction with the App, etc. will be presented, followed by a discussion on key topics to include when surveying respondents' on the user experience with cross-platform data collection tools.

The Mechanics of GPS Geo-Location for Mobile Devices: Their Potential for Measurement Error and Some Illustrative Data

Trashawna Boals, *Experian Marketing Services*; Max Kilger, *Experian Marketing Services*

As researchers continue their initial efforts to utilize mobile devices as survey research data collection instruments, more and more researchers will turn to taking advantage of some of the special features of these mobile devices as a part of their data collection process. In particular, the ability to pinpoint the exact geo-location of an individual may contribute information that will provide additional meaning to data already being collected by means of mobile devices. In this paper we examine the multiple technical mechanisms involved in geo-location using GPS services from mobile devices like Smartphones and tablets. In particular we look at the potential sources for and size of measurement error in each of these strategies — such as the “transporter effect” — due to a variety of factors. Finally, we examine some real-life geo-location data collected by a recent passive mobile measurement study to look for evidence of some of these potential measurement errors as well as provide the reader with some familiarity with mobile device-based geo-location data.

Assessing the Risk of Nonresponse Bias

Following up on Nonresponse Bias in the American Time Use Survey

Daniel G. Harwell, *National Center for Health Statistics*

Nonresponse can be a major issue when the topic of interest is directly related to response propensity. This issue is of particular concern for time use surveys, which measure how people spend their time. If individuals who are busier fail to respond to surveys due to a lack of time, this could lead to biased estimates. This has been a particular concern for the American Time Use Survey (ATUS), which has had a response rate below sixty percent since it began in 2003. Following up on previous research (Abraham, Maitland, and Bianchi 2006), this study examines the potential bias created by nonresponse in the 2011 ATUS, with particular emphasis on the theory that busier respondents are less likely to respond to the ATUS.

Multiple Approaches for Evaluating Nonresponse Bias in a Short-Field-Period Survey

Robyn Rapoport, *Social Science Research Solutions*; Paul J. Lavrakas, *Independent Consultant*; Eran Ben-Porath, *Social Science Research Solutions*; Melissa Herrmann, *Social Science Research Solutions*

High quality dual frame phone surveys typically employ several strategies to reduce non-response including: 1) making multiple call attempts to non-responsive numbers; 2) assigning highly trained interviewers to call back refusals in an effort to convert them into completed interviews; and 3) varying the time of day and day of the week that call attempts are made. Surveys that are fielded over short field periods (10 days or less) are limited in their ability to employ these types of approaches and may be more susceptible to the possible effects of non-response, including bias. This paper reports on a series of non-response bias studies embedded in a state-wide survey concerning voting that was fielded over a ten-day period in October 2012. Since the survey was expected to undergo intensive legal scrutiny regarding its validity in representing the population, the research team proactively incorporated four methods to investigate the presence of non-response bias. These included comparisons of data provided

by respondents who completed the survey in initial call attempts to those who completed in later call attempts and to those who had initially refused participation. The researchers also compared responders and non-responders using Census data associated with their local zip codes derived from telephone exchanges. For example, one variable found to significantly differentiate responders from non-responders was the percentage of the population in the zip code that was White. In order to gain insight into other types of potential non-response bias, the team reviewed information about refusals collected, in real-time, from the interviewers using a Refusal Report Form (cf. Lavrakas, 2010). Given that the survey industry is currently confronting diminishing response rates, particularly during short field-periods, pursuing a rigorous evaluation of the impact of non-response is imperative for increasing confidence in the validity of findings derived from this type of data collection.

An Evaluation of Alternative Indicators for the Risk of Nonresponse Bias for a Mail Survey With a Nonresponse Follow-Up

Sonja Ziniel, *Harvard Medical School; Boston Children's Hospital*; James Wagner, *University of Michigan*; Rebecca Hehn, *Boston Children's Hospital*; Robert Groves, *Georgetown University*; Ingrid Holm, *Boston Children's Hospital*

Recent research on nonresponse bias in surveys has included the development of alternative indicators for the representativeness of survey respondents, such as the R-indicators (Schouten, Cobben, and Bethlehem (2009)). Few empirical studies have been published and little is known about the usefulness of these indicators to detect nonresponse bias in survey statistics under different survey designs. This study evaluates these alternative indicators, including R-indicators, for a mail survey sent by Boston Children's Hospital to 7,000 parents whose children were recently seen at the hospital or any of the clinics affiliated with it. The survey focused on attitudes about participation in genetic biobanks and the return of genetic research results. Previous research indicated a high likelihood of nonresponse bias for a number of statistics from such a survey. After the initial survey and a reminder postcard were sent, we performed a nonresponse follow-up study for a random sample of the nonrespondents that included a shorter questionnaire and a \$2-bill as incentive. The highly detailed sampling frame included demographic and medical condition-related information on the children and parents. Exploratory analyses will compare nonresponse bias indicators across the two phases of the survey and assess their ability to detect nonresponse bias using data from the sampling frame as well as survey statistics.

The Effect of Survey Mode on Nonresponse Bias and Measurement Error: A Validation Approach

Antje Kirchner, *Institute for Employment Research*; Barbara Felderer, *Institute for Employment Research*

In order to obtain unbiased estimates from survey interviews, it is important that the data is of good quality: i.e. representativity of the survey respondents and variables that are free of measurement error. Using administrative records and survey data, the main questions we address concern the differential nonresponse bias between the telephone and the Web mode and whether these modes lead to differential measurement error. In an experimental setting we randomly assigned respondents to either phone (n=2,400) or Web mode (n=1,082). Because the sampled persons were selected from German administrative records, record data are available for all sample units to study the bias due to nonresponse and measurement error (e.g.

population means and regression parameters). Hence, we can assess the overall nonresponse bias of the estimates by comparing the statistics from both modes against the known population value. Similarly, for the respondents, we compare survey values and administrative records on individual level for selected variables and compute measurement error directly. First, based on administrative data for respondents and nonrespondents, our paper compares nonresponse bias for the above statistics in the single telephone mode to those obtained in the Web mode. Empirical analyses confirm a differential sample composition resulting in systematically different nonresponse bias between the two modes. Second, we assess the amount of measurement error for both modes. We conclude with a discussion of whether mode specific differences, with respect to nonresponse bias and measurement error bias, compensate or reinforce each other with respect to the total error.

Implications of Potential Nonresponse Bias

Ashton Jacobe, *Fors Marsh Group*

One of the challenges researchers face in data collection is achieving a representative sample of the population. There will usually be portions of the population who will not respond to the survey. If the people who do not respond are systematically different from those who respond, this introduces potential bias to the survey results. Therefore, survey nonresponse is a factor that plays a significant role in the composition of the resulting sample. Ideally, nonresponse bias would be measured by comparing the survey responses of those who did not respond to the survey to those who responded; however in most cases this is not possible. Hence the prevalence of potential nonresponse bias must be estimated. This paper discusses the nonresponse bias analysis for a sample of military recruiters used for a quality of life survey. The analysis used a two-stage process to approximate the differences in survey responses between sample members who did not complete the survey and those who completed it. The first stage compared demographic characteristics of respondents and nonrespondents. For those characteristics found to differ significantly between the two groups, responses to key survey items were analyzed in stage two to determine if the characteristics that differed between groups were related to responses to survey items. Results indicated that several key estimates showed potential bias; this response bias was primarily related to a few demographic characteristics (race/ethnicity, aptitude test score, and family characteristics). There were also differences in the amount of potential bias and drivers of the bias among the different Service subgroups. As a result of this analysis, the data were weighted to adjust for the potential biasing factors to ensure that all estimates better represent the full population of recruiters. This paper highlights the importance of determining the extent of potential nonresponse bias in survey data.

Culture and Survey Responses

Examining the Role of Culture in Answering Context-General and Context-Specific Survey Questions

Allyson L. Holbrook, *University of Illinois at Chicago*; Sharon Shavitt, *University of Illinois*; Timothy P. Johnson, *University of Illinois at Chicago*; Young I. Cho, *University of Wisconsin – Milwaukee*; Noel Chavez, *University of Illinois at Chicago*; Saul Weiner, *University of Illinois at Chicago*

Previous research has examined factors that influence responses to context-general (those that ask about opinions, behaviors, or beliefs that apply across contexts; e.g., general life

satisfaction) and context-specific (those that ask about opinions, behaviors, or beliefs that are limited to a particular context; e.g., satisfaction with one's work) questions. For example, the order of these questions (in particular part-whole pairs of questions) may influence the distribution of responses to the questions as well as the relationship between the questions (e.g., Schwarz, Strack, & Mai, 1991). We examine the role that culture may play in influencing responses to general and context-specific questions by examining several pairs of part-whole or specific-general question pairs (e.g., opposition to the death penalty for a specific crime and opposition to the death penalty in general) from a survey of non-Hispanic Whites, African-Americans, Mexican-Americans, and Korean-Americans. We assess the extent to which culture influences the process of answering these survey questions by examining order effects, respondent reactions to the questions (measured via coding of respondents' behaviors from recordings of the interviews), and paradata (e.g., response latencies) to test a number of hypotheses. For example, we expect that members of collectivistic cultures, who have been shown to think more contextually (Nisbett, 2003) will have more difficulty than individualists in answering the general items (which are less tied to context) but less difficulty with the specific items (which are more strongly tied to context). References: Nisbett Richard E. 2003. *The Geography of Thought*. New York: The Free Press. Schwarz, Norbert, Fritz Strack, and Hans-Peter Mai. 1991. "Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis." *Public Opinion Quarterly* 55(1):3-23.

Testing the Veracity of Self-Reported Religious Practice in the Muslim World **Philip Brenner, *University of Massachusetts Boston***

Survey findings suggest that predominantly Muslim countries are among the most religious in the world and validate commonly held, but overly simplistic, perceptions of Muslims as extremely and uniformly religious. Existing research has demonstrated that survey estimates can give a distorted view of the reality of levels of religious practice, however, it has thus far focused exclusively on traditionally Christian, advanced Western democracies. To address this oversight, the veracity of self-reported religious practice in the Muslim world is tested using Pakistan, the Palestinian Territories, and Turkey as cases for study. Comparing estimates of prayer from conventional surveys with those from time diaries, marginal rates of overreporting are estimated for each country by sex. The time use measure of prayer is then imputed for the conventional survey dataset to estimate overreporting at the respondent level and to predict overreporting using a measure of religious identity importance. Findings suggest that overreporting of prayer occurs in each country considered, although more consistently for women than men. These gender differences in data quality are discussed in terms of public/private religious practice. Moreover, religious identity importance is strongly correlated with overreporting of prayer, suggesting that a similar mechanism may promote the measurement error for overreported prayer in the Muslim world and overreported church attendance in the West.

Estaría Bien Si Le Hago Unas Pocas Preguntas En Ingles? An Experimental Investigation of Language Effects Among Bilingual Latinos

Nicole R. Buttermore, *Social Science Research Solutions*; Luis Tipan, *Social Science Research Solutions*; Mark Lopez, *Pew Hispanic Center*; David Dutwin, *Social Science Research Solutions*

A growing body of research has documented differences in survey results among Latino respondents interviewed in Spanish and those interviewed in English and demonstrated that the failure to offer interviewing in both languages introduces significant bias (e.g., Lee, et al., 2008; Dutwin et al., 2012). Such language effects might result either from: 1) cultural differences in attitudes, or; 2) differences in the meaning and interpretation of the questions depending on language (Perez, 2011). In an attempt to tease apart these explanations, we embedded an experiment in a national survey of Latino political attitudes and used random assignment to control for the effects of acculturation. During the middle of the interview, respondents who reported fluency in both English and Spanish were randomly assigned to either switch languages or continue in the language in which they began the interview. In line with previous research, the results demonstrated significant attitudinal differences between those answering in English versus Spanish, but there were also differences between those who did and did not change languages. Notably, of respondents who began the interview in Spanish, those who switched to English rated their financial situation as more favorable than did those who did not switch languages. Furthermore, respondents initially interviewed in English rated being successful in a high paying career as less important when responding in English, compared with those who switched to Spanish. These results suggest that the language in which a survey is administered has an important implications for the way in which respondents think about and respond to the questions.

Assessing the Validity and Reliability of Self-Reported Items on Likelihood of Migration

Sergio C. Wals, *University of Nebraska-Lincoln*; Alejandro Moreno, *Instituto Tecnológico Autónomo de México*

Immigrants provide a critical test to longstanding theories of attitude formation. These individuals, after all, import their political attitudes from their countries of origin to their new homes. Some of these attitudes and beliefs remain consequential for immigrants' civic lives whereas others are replaced through exposure to the new political system. With international migration flows on the rise and increased scholarly attention to immigrants' political attitudes, beliefs, and behaviors, social sciences are in need of valid and reliable survey instruments to study these hard-to-reach populations. Before and after migration panel data are ideal. Given budgetary and logistical constraints, however, before and after migration panel data are nearly impossible to collect. Building upon social science research on likelihood of migration, we develop a battery of items to assess the extent to which individuals from one country are likely to migrate to another one. We contend that a small battery of items can provide researchers with a theoretically-driven alternative to identify the most likely individuals to migrate from any given country to another in the near future, which in turn results in a cost-effective opportunity to gather pre-migration data on populations of interest. Over the past few decades, Mexico has provided the largest cohort of immigrants to the United States. Therefore, we tested our likelihood of migration battery on several nationally representative samples of Mexican citizens living in Mexico to identify those individuals most likely to migrate to the United States in the following two to three years. Our data analyses assess the validity and reliability of our survey

items. The data collection took place from 2007 to 2012. Our empirical findings strongly suggest that our likelihood of migration battery is a viable and efficient option to collect valid and reliable pre-migration data on populations of interest.

A Cross-Cultural Study on Daily Experience of Depression Between Countries in the Sahel Region and Western Asia

Jinyoung Lee, *University of Nebraska - Lincoln*

According to the World Health Organization's World Mental Health Survey Initiative (2011), richer countries have higher depression rates. Contrary to this finding, the Gallup World Poll has shown that countries suffering from extreme poverty and/or wars have the highest depression rates: people in Ethiopia (43.4%), the Palestinian Territories (30.8%), Yemen (28.2%), and Iraq (26.3%) report the highest rates of depression, while less than 10% of people feel depressed in most of the developed countries such as Denmark (4.2%), Netherlands (4.8%), Switzerland (4.8%), and Sweden (4.9%). The remarkable difference between the two results implies that studies on depression should pay attention to the causes of depression differentiated by the social context of each country. This study examines daily experience of depression based on the Gallup World Poll, a multinational probability-based survey. This dataset does not distinguish clinical depression from momentarily feeling down. This study focuses on the socio-political conditions that contribute to the public's feeling of depression. As Diener et al. (2003) noted, comparative studies on emotional and cognitive aspects are complicated because cultural variables as well as personalities yield differences in mean level of individual evaluation of life between countries. Considering this complexity, this study traces fluctuations in depression rates within countries in the Sahel region or Western Asia, rather than simply stating that the public of which country are more depressed. A preliminary analysis indicates that depression rates might reflect major socio-political events such as famine and violence against the citizens, that is, the depression rates of a country are often influenced by major social events. Further, different factors such as extreme poverty and fear result in different levels of depression rates between countries. The result confirms that cross-cultural studies on depression should meticulously take the unique social context in each country into consideration.

Friday, May 17
10:00 a.m. – 11:30 p.m.
AAPOR Concurrent Session D

Probability and Non-Probability Samples in Internet Surveys

Understanding Bias in Probability and Non-Probability Samples of a Rare Population

John Boyle, *ICF International*; Sarah Ball, *Abt Associates*; Helen Ding, *Chenega Government Consulting LLC*; Gary L. Euler, *NCIRD, Centers for Disease Control and Prevention*; Stacie Greby, *Centers for Disease Control and Prevention*; Faith Lewis, *Abt SRBI*

Although probability samples are the preferred source for national data measures, non-response issues in probability samples have a substantial impact on the cost of surveys among rare populations. In some rare populations, non-probability samples can produce credible results that would not otherwise be possible to obtain. CDC conducts surveys of flu vaccination attitudes and usage among pregnant women in order to monitor health risks in this vulnerable population. Pregnant women account for 1% of the adult U.S. population. The National Health Interview Survey (NHIS) identifies pregnant women in its sample; however, the sample size is small and the data are not available by the start of the next flu season. During the 2010-11 flu season, CDC conducted nearly 1,500 interviews of pregnant women in the fall and 2,000 interviews in the spring using a large national Internet panel. The fall Internet panel surveys are launched on or around November 1 and published in early December to provide an early season estimate of flu vaccination among pregnant women. The final season estimates measured by the spring Internet panel surveys are published at the start of the following flu vaccination season. Complete data for pregnant women in the NHIS are not available until near the end of the following flu vaccination season. In this paper we examine the characteristics of pregnant women in the Internet sample and compare them to the NHIS to better understand potential sources of error in both probability and non-probability samples, with consideration for reasons to choose between a probability and non-probability sample for generating rapid data to assess public health programs.

A Comparison of Results from Dual Frame RDD Telephone Surveys and Google Consumer Surveys

Scott Keeter, *Pew Research Center*; Leah Christian, *Pew Research Center*; Danielle Gewurz, *Pew Research Center*; Michael Dimock, *Pew Research Center*; Rob Suls, *Pew Research Center*; Jon Sadow, *Google*; Paul McDonald, *Google*; Brett Slatkin, *Google*; Matt Mohebbi, *Google*

The growth in Internet use has led to the development of new techniques for conducting social research and measuring people's behavior and opinion while they are online. One such tool, Google Consumer Surveys, interviews a sample of Internet users from a diverse group of about 80 publisher sites that allow Google to ask one or two questions of selected visitors as they seek to view content on the site. Google's approach results in a nonprobability sample of Internet users, but is distinct from opt-in surveys in that respondents cannot self-select into the

survey. It is also different from Internet panels that respondents join for an extended period of time. This paper will summarize a year-long evaluation of how results from Google Consumer Surveys compare with those from dual frame RDD telephone surveys. Specifically we compare Google and telephone survey estimates across a wide range of political attitudes and behavior, domestic and foreign policy opinions, technology use and civic and political engagement. We also examine how the demographic composition of the samples compares with that of Internet users in both telephone surveys and the Current Population Survey. In the initial six months of evaluation, the median difference in point estimates across substantive and demographic 48 questions tested was 3 percentage points; the mean difference was 6 percentage points. The paper will offer guidance on how Google Consumer Surveys can be used for different applications of survey research.

A Comparison of a Mailed-in Probability Sample Survey and a Non-Probability Internet Panel Survey for Assessing Self-Reported Influenza Vaccination Levels Among Pregnant Women

James Singleton, *Centers for Disease Control and Prevention*; Helen Ding, *Chenega Government Consulting LLC*; Stacie Greby, *Centers for Disease Control and Prevention*; Gary L. Euler, *NCRID, Centers for Disease Control and Prevention*; Indu B. Ahluwalia, *NCCDPHP, Centers for Disease Control and Prevention*; John Boyle, *ICF International*

The Centers for Disease Control and Prevention (CDC) conducted an opt-in Internet panel survey (IPS) to provide timely estimates of mid-season and end-of-season influenza vaccination coverage among pregnant women. We used the Pregnancy Risk Assessment Monitoring System (PRAMS), a stratified probability sampling survey, to assess the representativeness of the pregnant women sample and the validity of influenza vaccination coverage from IPS. The IPS is an “opt-in” survey for women who were pregnant anytime since August 2010-April 2011. PRAMS is an ongoing population-based surveillance system collecting data on maternal experience and behaviors before, during and shortly after pregnancy among women delivering a live-born infant. For both surveys, we limited the analysis to women pregnant during the peak flu vaccination period (October 2010-January 2011) residing in 18 states with completed data and compared final weighted distributions of demographic characteristics and influenza vaccination coverage and un-weighted IPS vs. base-weighted PRAMS (accounting for probability of selection) distributions of demographic characteristics. Compared to PRAMS, IPS respondents had similar age and marital status distributions, were more likely to be white (68.1% vs.58.0%), less likely to be Hispanic (10.7% vs.16.5%) or other racial/ethnic groups (6.0% vs.12.0%) before final weighting. The higher percentage of women with college and above education from IPS (44.0% vs.34.6%) persisted after final weighting (43.6% vs.31.9%). Overall influenza vaccination coverage from both surveys was similar (50.2% vs.49.2%) and the estimates by subgroups were similar ($\pm < 5\%$) except by race/ethnicity ($\pm > 5\%$). While neither survey provides a standard measure of flu vaccination among pregnant women, IPS is able to provide a similar vaccination coverage estimates among pregnant women within the flu season for rapid response, while PRAMS provides detailed state level data for longer term planning. Both surveys will be continued to assess immunization programs and to ensure valid, timely data available for decision making.

Probability vs. Non-Probability Samples: A Comparison of Five Surveys

Johan Martinsson, *University of Gothenburg*; Stefan Dahlberg, *University of Gothenburg*; Sebastian Lundmark, *University of Gothenburg*

Commercial Internet panels based on non-probability samples have begun to be widely used, also by academic researchers. But is the quality of such data comparable to that of probability based samples? Previous studies addressing this issue have mainly focused on the U.S., while this study compares the quality of such Internet panels in different context: Sweden. Sweden differs from the U.S. for example by having had very high Internet coverage for a long time, having smaller socio-economic differences and by having a complete population register that can be used for random samples. Two non-probability based panels are compared with two probability based panels and a benchmark telephone survey. Demographics are compared to government records, and attitudes are compared to benchmark studies of high quality and high response rate. In order to allow comparisons five surveys with comparable questions were run at the same time. Three of the Internet panels provided professional post-stratification weights, which allow us to compare the accuracy both with and without weights. In contrast to previous studies, the results indicate a surprising similarity in terms of accuracy between probability panels and non-probability panels. The reasons for this deviating result and differences between the United States and Sweden are discussed.

Modeling a Probability Sample? An Evaluation of Sample Matching for an Internet Measurement Panel

Lukasz Chmura, *The Nielsen Company*; Douglas Rivers, *YouGov*; Delia Bailey, *YouGov*; Christine Pierce, *The Nielsen Company*; Scott Bell, *The Nielsen Company*

The past several years have seen an increase in the usage of samples from opt-in panels. While these samples are relatively inexpensive compared to more traditional sample designs, they are subject to unknown biases. A sample matching approach was evaluated as a means of controlling and reducing this bias. Sample matching allows us to select a subset of the opt-in sample that is as similar as possible, at a sample unit level, to a probability sample based on variables common to both. Assuming selection into the sample is independent of the survey variables conditional upon the matching variables, the matched sample will produce consistent results. With assistance from YouGov, Nielsen evaluated the effectiveness of the sample matching process to measure Internet behavioral metrics, including individual site visitation and Internet usage levels. The American Community Survey (a respondent level, publicly available probability sample) was used as the target sample for matching, and a matched sample was selected from the non-probability component of the Nielsen Netview panel (Nielsen's Internet audience measurement service, comprised of both probability and non-probability components). The resulting matched sample was compared to the probability portion of the Netview panel. The results are encouraging, showing significant reductions in bias among the matched sample. Specifically, for the Internet behavioral metrics considered, the matched sample generally produced smaller differences from the probability panel than the full opt-in sample, even after standard post-stratification weighting was applied. The robustness of the method was also evaluated, with the matched samples producing relatively stable estimates. While additional research is necessary to fully optimize this new methodology, the results so far show promise in producing quality results from an opt-in sample.

Question Construction and Data Quality

Impact of Filter Questions on Estimates of Media Consumption

Curtiss Cobb, *GfK Knowledge Networks*; Danell Godinez, *GfK Knowledge Networks*; Randall Thomas, *GfK Knowledge Networks*; Julian Baim, *GfK-MRI*; Risa Becker, *GfK-MRI*

A key choice in the design of Web surveys is whether to avoid posing questions to respondents that do not apply to them by first asking filter questions. In research on filter questions, there is some indication that a dichotomous “yes” or “no” response will yield a lower proportion of self-reported occurrences of behaviors or attitudes than a multi-category scale. For example, in a number of studies measuring attitudes (e.g. ‘concern’) or self-reports of ‘crime,’ multi-category formats have been associated with higher self-reported incidence or attitudes than conditions that filter with yes-no formats (Herrmann, et al., 1998; Hippler and Schwarz, 1989; Knäuper, 1998; Sterngold, et al., 1994). These findings are at odds with the cognitive processes that survey researchers and psychologists believe that respondents use to answer questions. It is believed that respondents first determine whether an incident or attitude occurred and before trying to map it onto the provided multi-category response scale. This study extends the research on filter questions by examining their use to measure media consumption, particularly newspapers readership, radio listening and television viewing. Using a split ballot design on a representative sample of 1,000 adults, we randomly assigned half the sample to report their media consumption over a period of time using a multi-category response, while the other half of the sample were first asked a filter question before receiving the multi-category response if eligible. Preliminary findings show that respondents receiving the multi-category response reported more media consumption than those receiving the filter questions. Additional analysis will explore differences along demographic lines and seek to relate the findings of non-media use to satisficing behavior in other parts of the survey instrument.

Response Format Effects in the Measurement of Employment

Sergei Rodkin, *Gfk Custom Research, LLC*; Randall K. Thomas, *Gfk Custom Research, LLC*; Stefan Subias, *Gfk Custom Research, LLC*; Carolyn Chu, *Gfk Custom Research, LLC*

Accurate measurement of employment is essential to track employment trends in a nation, with the information used to determine the effectiveness of a variety of private and governmental programs designed to increase employment. Some have noted discrepancies in estimated employment numbers between the Census and the CPS (Census typically has a lower count of employed people), most often attributed to differences in interviewing mode, time frame reference, or sampling frame. Many researchers using paper-pencil or Web-based questionnaires present a multiple response question (‘Select all that apply’) to assess employment. However, in a telephone interview, employment is often asked through a series of yes-no questions, with the interviewer requesting a ‘yes’ or ‘no’ response for each item presented in sequence (cf. Smyth, Christian, and Dillman, 2008, POQ). In research with self-administered questionnaires, the Yes-No Grid format has been found to yield a higher level of endorsement than the Multiple Response format in self-administered surveys (Smyth, Dillman, Christian, and Stern, 2006, POQ; Thomas and Klein, 2006, JOS). This paper reports on two studies – Study 1 was a Web-based study that was conducted across 24 monthly waves with over 60,000 respondents (18 or older) using an opt-in non-probability panel, balanced demographically for age, sex, region, education, and income. Respondents were randomly assigned to one of the 3 employment response scale formats: Multiple Response Format

(MRF); Yes-No Grid (YNG for employment); Single Response Format (SRF). Study 2 was a Web-based study with over 2700 respondents using a probability-recruited panel (GfK-Knowledge Networks) with the same conditions used in Study 1. In both studies, endorsement of every category was higher with the YNG and lowest with the SRF. We will also describe how these results are related to trend changes across quarters and how they are related to other work-related variables, including hours worked/week.

Grouped Versus Interleafed Questions and Specific Versus Global Questions to Improve Accuracy of the Census Questionnaire

Emily Geisen, *RTI International*; Murrey Olmsted, *RTI International*; Jennifer H. Childs, *U.S. Census Bureau*

To reduce duplication or misreporting on the census, the U.S. Census Bureau includes questions asking respondents about alternate addresses where household members sometimes live or stay. However, recent studies based on the 2010 census found evidence of underreporting these alternate addresses. To improve enumeration and reduce costs associated with conducting follow-up interviews, the Census Bureau is exploring the use of computer-assisted interviewing (CAI) for the 2020 census. The use of CAI for the 2020 Census allows us to explore the use of two different design methods to encourage reporting of alternate addresses: (1) grouped questions versus interleaved questions, and (2) specific versus global questions. Research has shown that asking a group of yes/no filter questions before asking detailed follow-up questions can elicit more “yes” responses compared to interleaved questions, where follow-up questions come immediately after the filter question. (Kreuter, McCulloch, Presser, & Tourangeau, 2011). This idea is incorporated into the 2020 Census by asking respondents a series of yes/no questions about whether household members live or stay somewhere else before asking for the more detailed address information, which respondents may be reluctant to provide. Furthermore, research shows that asking global questions (i.e., asking about the household collectively) elicits less detail from respondents than asking specific questions (i.e., asking about each household member individually). However, using specific questions can lead to lengthier surveys and increased respondent burden. In this paper, we examined the qualitative results of these two design methods compared to a control using two rounds of cognitive and usability testing with approximately 100 participants. We examined which methods resulted in higher reporting of alternate addresses, and higher reporting of household members with alternate addresses. In addition, we investigated which method was associated with more accurate reporting overall, the least number of user errors, and the lower respondent burden.

Minor Design Changes With Major Impacts: Testing Explicit Versus Implicit Don't Know and Refused Response Options in Audio Computer-Assisted Self Interviewing

James M. Dahlhamer, *National Center for Health Statistics*; Adena Galinsky, *National Center for Health Statistics*; Sarah Joestl, *National Center for Health Statistics*; Marcie Cynamon, *National Center for Health Statistics*; Jennifer Madans, *National Center for Health Statistics*; Virginia Cain, *National Center for Health Statistics*

An ongoing debate among survey researchers focuses on the provision, or not, of an explicit don't know response option for questions in self-administered surveys. Some argue that offering an explicit don't know option invites satisficing, an “easy way out,” resulting in elevated item

nonresponse. Others counter, arguing that the exclusion of an explicit option represents a form of coercion, forcing respondents to answer when the required knowledge/experience/opinion does not exist. To inform this debate, we utilize data from two 2012 field tests evaluating the feasibility of audio computer-assisted self-interviewing (ACASI) in the National Health Interview Survey. In the first test, questions on sexual identity, mental and financial health, sleep, and HIV testing were administered via ACASI to 535 adults. Don't know and refused options were provided with each question, and respondents could advance without answering ("explicit" approach). The second field test involved a split-ballot experiment in which 3,215 adults were assigned to receive questions using ACASI and 2,237 using computer-assisted personal interviewing (CAPI). For ACASI, explicit don't know and refused options were eliminated and a follow-up item (response options: return to question, don't know, refused) was presented when a respondent advanced without answering ("implicit" approach). We assessed the ACASI design changes by comparing item nonresponse rates between ACASI cases from the two field tests. Where significant bivariate results emerged, the impact of screen design was tested in a multivariate setting, controlling for sociodemographic characteristics such as age, sex, and education. We also assessed mode differences by comparing item nonresponse rates between CAPI and ACASI cases from the second test. Here again, significant bivariate results were followed by multivariate analyses controlling for sociodemographic measures. Preliminary results suggest a considerable advantage to the implicit approach. We conclude by discussing the implications of our results for self-administered questionnaire design.

Seymour Sudman Student Paper Award Winner

Measure for Measure: An Experimental Test of Online Political Media Exposure
Andrew Guess, *Columbia University*

It is well known that existing measures of self-reported political media exposure are potentially unreliable. Various studies have explored the causes of such measurement error, such as social desirability bias, and have tested proxies, such as political knowledge. However, lacking an objective baseline, investigations of this sort still rely solely on survey responses. By focusing specifically on recent Internet activity, this paper's methodology estimates individuals' actual consumption of political media. Using an experiment embedded within an online survey, I test two different measures of media exposure and compare them to the estimated actual exposure. I find that open-ended prompts produce generally more accurate measures of recent exposure to online media compared to multiple-choice questions offering a list of different political news outlets, which tend to produce overreporting.

Interviewing Methods and Survey Outcomes

Rapport, Sensitivity, and Proxy Reporting: Questions About End-of-Life Planning and Interviewer-Respondent Interaction

Dana Garbarski, *University of Wisconsin-Madison*; Nora Cate Schaeffer, *University of Wisconsin-Madison*; Jennifer Dykema, *University of Wisconsin Survey Center*

"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 feature of the interaction, referring to a situated sense of affiliation between interactional partners, comfort, willingness to disclose, motivation to please, empathy, or sharing (Goudy and Potter, 1976). Rapport may benefit response quality by increasing respondent motivation, but could also harm data quality.

Questions about one's end-of-life treatment 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. We propose to consider the various meanings and dimensions of rapport and consider what their interactional expressions might be both for the interviewer and the respondent in this context. This study examines transcripts of the end-of-life section of the 2004 wave of the Wisconsin Longitudinal Study in order to examine the conversational practices through which interviewers and respondents negotiate sensitive topics and answers in terms of the signals that respondents give about sensitive answers and how interviewers signal these questions are delicate, ask these questions, and follow-up respondent answers. A coding scheme is developed to examine the features of the interviewer-respondent interaction, including behaviors associated with rapport, sensitivity, and motivation as outlined above, as well as behaviors previously identified as indicating potential problems in the response process such as markers of uncertainty (see, e.g., Garbarski, Schaeffer, and Dykema, 2011). These coded features of interviewer-respondent interaction will be examined for their associations with two criteria: participation in a subsequent wave of the Wisconsin Longitudinal Study, and, for respondents who are married, concordance with spouses in reports of spouses' end of life treatment preferences.

Measuring Conversational Interviewing and Its Impact on Data Quality in the American Time Use Survey

Scott Fricker, *U.S. Bureau of Labor Statistics*; Morgan Earp, *U.S. Bureau of Labor Statistics*; Jennifer Edgar, *U.S. Bureau of Labor Statistics*; Polly Phipps, *U.S. Bureau of Labor Statistics*; Stephanie Denton, *U.S. Bureau of Labor Statistics*

In the American Time Use Survey (ATUS), interviewers use a set of scripted open-ended questions to walk respondents chronologically through the prior 24-hour day, collecting activities and details about each activity reported. The interview is designed to be administered using conversational interviewing, a method thought to put the respondent at ease and provide interviewers with the freedom to collect data in the best possible way. Conversational interviewing is hypothesized to improve respondent understanding of questions and concepts as interviewer and respondent converse and collaborate on meaning. In the ATUS, conversational interviewing also is thought to improve recall by allowing interviewers to ask open-ended questions to assist respondents in reconstructing their day in a way that is meaningful to them rather than following a set script and sequence. Previous research has explored the use of conversational interviewing in the ATUS and found that although some conversational interviewing methods are used, they are not used consistently across all interviewers, respondents, or even within an interview. The impact of conversational interviewing techniques on data quality also was found to be inconsistent. In this paper, we use 100 behavior-coded transcripts of ATUS interviews to further explore the use and scope of conversational interviewing and the impact on data quality. We identify important components of conversational interviewing, based on interviewer behaviors and respondent-interviewer interactions, and develop a scale to measure how conversational the interview is. New measures of data quality, including the adequacy of respondent answers, number and type of respondent activities, missing activities, and interview length are explored and multivariate analysis techniques are utilized to better understand the complex relationship between interviewer and respondent behaviors, as well as the quality of the data collected.

Predicting the Occurrence of Respondent Retrieval Strategies in Calendar Interviewing: The Quality of Retrospective Reports

Robert F. Belli, *University of Nebraska – Lincoln*; L.D. Miller, *University of Nebraska – Lincoln*; Leen Kiat Soh, *University of Nebraska – Lincoln*; Tarek Al Baghal, *University of Nebraska – Lincoln*

Calendar based survey interviewing methods have been predicted to enhance the quality of retrospective reports by encouraging the use of thematic and temporal retrieval cues that reside in autobiographical memory. These cues—measured as observable verbal behaviors—exist as parallel (using a contemporaneous event from the respondent's past to cue an event in a different theme) and sequential (using an event to remember what happened earlier or later within the same theme) interviewer probes, and respondent parallel and sequential retrieval strategies. Previous research has shown that retrieval behaviors are associated with better retrospective reporting data quality when the respondents' histories are more complex. The current study focused on discovering patterns of interviewer verbal behaviors that predict the occurrence of respondent parallel retrieval strategies, and whether these patterns are associated with data quality. Data are derived from the interviews of 153 respondents of the Panel Study of Income Dynamics (PSID) who were interviewed about their life course histories. For every respondent turn of speech, the occurrence or nonoccurrence of a respondent parallel retrieval strategy was evenly sampled. Verbal behaviors of immediately preceding interviewer and respondent turns of speech were assessed in terms of their co-occurrence with parallel retrievals using a decision tree data mining algorithm called C4.5. We discover seven patterns of preceding behaviors that have the most impact on encouraging respondent parallel retrieval strategies. We assessed the association between the occurrences per interview of each of these patterns on response accuracy in reports of employment as determined by comparing calendar responses with responses collected in prior waves of the PSID. Interviewer sequential probing, when followed by a respondent parallel, was associated with greater accuracy, but interviewer parallel probing was not. For some patterns that involved sequential probing, greater accuracy was only observed with respondents who had complicated employment histories.

Linking Interview Context, Interviewer Behavior and Data Quality

Aaron Maitland, *Westat*; Wendy Hicks, *Westat*

Interviewers play an important role in gaining the cooperation of survey respondents and administering questions. Several studies have explored the relationship between interviewer behavior and different sources of survey error. But there is little known about the mechanism in which interviewers' affect error. A study by Olson and Peytchev (2007) adds some insight into the interviewers' effect. The authors found that as interviewers conduct more interviews, the length of the interview decreases and the interviewers perceive the respondents as less interested. While we would anticipate that interviewers improve their skill in navigating and administering an instrument over repeated administrations, the change in interviewers' perception of respondents' interest in the study may not be independent of the faster administration and may actually be more reflective of the interviewers' own attitude. We build on these findings and make use of Computer Audio Recorded Interviewing (CARI) and coding analysis to further understand the mechanism in which interviewers' behaviors may affect error. Using the National Health and Aging Trends Study (NHATS), we link behavior coding analysis with contact history data and interviewer characteristics to create a context in which we examine the relationship between interviewer behavior and data quality. In the analysis, we construct a 'case difficulty' variable based on the contact history data and compare interviewer behaviors

between the more difficult and less difficult cases. In addition, we account for interviewer productivity as a variable related to interviewer behaviors. In a preliminary analysis, we found that interviewers differ in how well they follow the standardized interviewing protocol between difficult and less difficult cases, depending on their overall productivity. In this paper, we look at whether there are differences in data quality as measured by item nonresponse, interview length and the consistency of survey responses when interviewers' deviate from protocol.

Hello? Is Better Than Hello: Effects of Greetings on Participation in Survey Invitations

José R. Benkí, *University of Michigan*; Jessica Broome, *University of Michigan*; Frederick Conrad, *University of Michigan*; Robert Groves, *Georgetown University*; Frauke Kreuter, *University of Maryland JPSM & IAB*

Potential respondents to telephone survey interviews rapidly decide whether or not to participate, with most refusals occurring within 30 seconds of answering the phone. Given the speed of this decision, it is likely that the initial verbal interactions between the interviewer and the "answerer" play an important role in the answerer's decision to participate. The present study focuses on the acoustic properties of "Hello" greetings by interviewers and telephone answerers at the beginning of survey invitations, and the relationship of these properties to the outcome of specific telephone survey invitations: agree-to-participate, scheduled-callback, and refusal. These relationships are explored in a corpus of 1380 audio-recorded contacts, sampled from five studies conducted by the University of Michigan Survey Research Center. Half of the contacts contain "hello" greetings suitable for acoustic analysis, including pitch measurement. Following Schegloff (1998), who documents how high-pitched greetings in telephone conversations signal enthusiasm, recognition, and friendliness, we hypothesize that contacts containing high-pitched "hello"s are more likely to lead to agreement or scheduled-callback instead of refusal. Greetings resulting in refusal contained average pitch rises of 18% above baseline pitch level for both answerers and interviewers. Contacts resulting in agreement or scheduled-callback contained greetings with higher pitch rises, 22% for answerers and 26% for interviewers. The significantly higher interviewer pitch rises in nonrefusals suggests that the positive attributes conveyed through high-pitched greetings promote participation. A second analysis considered the interaction between answerer and the interviewer greeting intonation. Consistent with the previous result, the highest rate of agreement occurs for contacts in which both actors produce a greater-than-average pitch rise. The lowest rate of agreement occurs for contacts in which the answerer greeting contains a pitch rise, but the subsequent interviewer greeting had flat intonation, suggesting that interviewer failure to reciprocate an enthusiastic and friendly greeting can be particularly harmful to participation.

Decision-Making in the 2012 Election

Validating Likely Voter Measures in 2012 Pre-Election Polling

Jocelyn Kiley, *Pew Research Center*; Scott Keeter, *Pew Research Center*; Matt Frei, *Pew Research Center*; Seth Motel, *Pew Research Center*; Leah M. Christian, *Pew Research Center*; Michael Dimock, *Pew Research Center*; Michael P. McDonald, *George Mason University*; Matthew Berent, *Matt Berent Consulting*; Jon Krosnick, *Stanford University*

One of the key challenges in pre-election surveys is determining the likely electorate. Substantial research has shown that people over-report their intention to vote (Holbrook and

Krosnick 2010, McDonald 2011) so pollsters have developed various methods of identifying which survey respondents are likely to vote and which are not. However, the accuracy of these methods have rarely been validated using actual voter records, aside from Paul Perry's path breaking work in this area in the 1960s, and a Pew Research Center study in a local mayoral race in the late 1990s (Perry 1970 and Dimock, et al. 2001). This paper will present preliminary analysis on the effectiveness of various likely voter measures using Pew Research's 2012 final pre-election poll of 3,815 adults and a post-election survey of a sample of registered voters from the same survey, in which voter records will be used to distinguish between voters and nonvoters. The analysis will explore how various survey questions designed to identify likely voters correlate with actual turnout. The paper also will explore the relatively new issue of how best to handle respondents who say they have "already voted", who constituted only a very small proportion in past years. In particular, it will examine the extent to which over-reporting of voting occurs among those who report having already voted.

The Impact of the Presidential Debates on Undecided and Persuadable Voters
Curtiss Cobb, *GfK Knowledge Networks*; Charles DiSogra, *Abt SRBI*; Jordon Peugh, *GfK Knowledge Networks*; Sarah Dutton, *CBS*; Anthony Salvanto, *CBS*

While politicians and pundits heralded Gov. Mitt Romney's performance in the first debate of the 2012 presidential campaign as a game-changer, and included the subsequent narrowing of support between the candidates in public opinion polls as evidence, political scientists were warning that it was all likely hype. Past research on debates have found little in the way of direct effects on candidate support and instead lead to partisan reinforcement (Hyllygus & Simon 2003; Kenski & Jamieson 2006). Moreover, debate "effects" are in part mediated through the post-debate political conversation (Brubaker & Hanson 2009). This two-wave study examines how the 2012 presidential debates and the subsequent post-debate conversation altered undecided and persuadable voters' perceptions of the candidates and their ultimate vote choice. Using GfK's probability based Internet panel, KnowledgePanel®, a group of undecided and persuadable voters were identified prior to each presidential debate and asked to complete a post-debate questionnaire in the hour immediately following the debate. Respondents were asked about their impressions of the debate performance of each candidate and for whom they planned to vote and why. Respondents were then re-interviewed on Election Day, along with decided voters and undecided voters that failed to watch the debates. They were asked again about their impressions of the candidates' debate performances and who they voted for and why, along with questions about media consumption, political interest and political knowledge. Differences between the three groups are being analyzed. The analysis will show: (1) whether undecided debate watchers votes differed from undecided non-watchers and already-decided voters; (2) whether a re-evaluation of debate performance occurred in the days between the debate and the election, and if so whether it relates to media consumption; (3) and whether political interest and political knowledge are moderating variables.

The RAND Continuous 2012 Presidential Election Poll
Tania L. Gutsche, *RAND Corporation*; Arie Kapteyn, *RAND Corporation*; Erik Meijer, *RAND Corporation*; Bas Weerman, *RAND Corporation*

The RAND Continuous 2012 Presidential Election Poll (CPEP) was conducted within the American Life Panel, which is an Internet panel recruited through traditional probability sampling to ensure representativeness. The CPEP differs from other polls in that it asks the same respondents repeatedly about their voting preferences. Thus, it leads to more stable outcomes

and changes are due to individuals' changing their minds and not due to random sampling fluctuations. The CPEP is also different because it asks respondents to state their preferences for a candidate and the likelihood that they will vote in probabilistic terms (percent chance). Moreover, we asked the panel members after the election whether they had voted and who they had voted for, so we can study the predictive power both within sample and out of sample (the national results). The CPEP appears to have predicted well. Our final prediction of the difference in popular vote between Obama and Romney differed less than .7 percentage point from the final tally. The probabilistic questions, even months before the election, were strongly related to individuals' actual voting behavior. Our approach allows us to gain insights in stability of voting preferences and the effect of events on individual preferences; for example, we see that changes in intention to vote play an important role in predicted vote shares for the candidates, while various shifts can be related clearly to major events. The American Life Panel has a wealth of background characteristics which can be related to voting preferences.

Survey Research as a Campaign Tool: Turnout Effects of Survey Respondents **David M. Margolis, *Greenberg Quinlan Rosner Research***

Researchers and political operatives alike are both concerned with determining which campaign methods and tactics are best for boosting voter turnout. Whether advertising, door-to-door canvassing, telephone calls, or direct mail, many methods of traditional voter contact have been tested in growing experimental literature on the effectiveness of voter mobilization efforts. In the case of political campaigns, one of the most common methods of conducting research on voter behavior adopts the same mode of one of the common methods of promoting voter mobilization: telephone surveys. Given the similarity of the mode (in fact, many voter mobilization call scripts adopt a format of an opinion survey), the effect that responding to a public opinion survey has on the likelihood that the respondent will turn out to vote should be evaluated. The author will implement a vote propensity score matching process to evaluate the effect that taking a political survey has on voter turnout likelihood. This quasi-experimental design will compare similar individuals in the treatment group (survey respondents) with non-treated individuals (sample members who did not receive a contact), and assess the voting behavior of both groups. Similar individuals will be identified using a listed sample where all potential respondents were assigned a modeled vote propensity score, and the average treatment effect can be analyzed with a paired difference test.

The Influence of Social Desirability in the Rise of Political Independents **Samara Klar, *Northwestern University*; Yanna Krupnikov, *Northwestern University***

Over the last several decades, survey researchers have seen a consistent change in political party identification among the American public. Today, when asked by researchers and pollsters, a plurality of Americans identifies themselves as Independent, as opposed to committing to one of the two parties. This evident detachment from partisanship brings with it scholarly concerns for citizens' engagement with the party system. Using a series of survey experiments, we demonstrate that these shifts are not, in fact, indicative of a genuine decline of partisanship, but rather a function of heretofore undetected social desirability pressures. We show that this tendency to identify as Independent is particularly likely to be triggered by media coverage focusing on the importance of undecided and Independent voters, as well as coverage emphasizing bickering between the two parties. Our two-wave panel experiment allows us to measure individual partisan preferences prior to stimulus exposure, strengthening the case that media coverage increases the social desirability of identifying as an Independent. Our follow-up

study demonstrates the implications of this finding for both survey research and for political participation more broadly.

Questionnaire Translation: Janet Harkness' Contributions, Legacy, and Beyond

This is one of two sessions to review and honor the contributions of Janet Harkness to the field of survey translation, adaptation, and questionnaire development in multilingual and multicultural surveys. All of Harkness' contributions have advanced our understanding of language and cultural issues in conducting surveys across language, cultures and regions. Where is the impact of her work most visible and what significance does this have on measurement in multicultural, multilingual surveys? How has this led to improvements and what are the areas where her legacy has the potential to largely improve procedures? Villar and Schoua-Glusberg's paper will focus on this overview of Harkness' work and contributions. Methods to improve equivalence in measurement tools across languages/cultures was a central focus of Harkness' work. When is translation not enough to produce equivalent measures for comparative studies, and adaptation is required? What are the limits of adaptation to still maintain comparability? Behr addresses these issues and focuses on the need to strive for a common understanding of adaptation and its applicability. Harkness strived to advance the field of survey translation by conducting and promoting basic methodological research to uncover the strengths and weaknesses of different approaches to translation and translation assessment. The other two papers in this session are examples of such research. Dörner examines the effects of advance translation of selected items in the ESS and how it allowed to uncover issues in the original items, leading to changes in the English version. This is precisely the flexibility that Harkness hoped questionnaires would have to adopt changes based on the need to reach better crossnational equivalence. In Schoua-Glusberg's paper, another experiment of a survey measure translation into Polish will provide evidence on what backtranslation and the committee approach can each contribute to translation assessment and what are their strengths and weaknesses.

Overview of Janet Harkness' Work and Contributions to the Field: Where Did She Lead Us To and Where We Are Now

Ana Villar, *Research Fellow*; Alisú Schoua-Glusberg, *Research Support Services*

This paper aims to present and evaluate the contributions of Janet Harkness to the field of multinational, multilingual, and multicultural (3M) surveys. Her contributions have advanced our understanding of language and cultural issues in conducting 3M surveys and her legacy has the potential to largely improve procedures used to implement them. We will start by reviewing the main areas in which the impact of Harkness' work is visible and the significance this has on measurement and comparability. As leader of the ESS Translation Task Force and of the Translation and Questionnaire Design Group of the ISSP, she helped set up translation procedures, stressing the importance of incorporating cross-cultural input at the questionnaire design stage to increase the chances of translations resulting in comparable instruments. She proposed a number of techniques (e.g., advance translation) and translation models (TRAPD) that are currently used in important international projects as well as large national projects with federal and international funding. She advocated the use of pretesting methods to assess translation and challenged existing views on answer scale construction. Some of Harkness' contributions, however, have not yet had the impact on survey implementation that they may

one day have. Many studies still follow a strategy where the source questionnaire is developed and finalized without taking into consideration input from the other cultures or languages that questionnaires will be translated to. Even worse, in smaller national projects that add a minority language as a last step before fieldwork, the resulting translation often is not done in time to go through appropriate assessment procedures, and thus its quality is very often unacceptable. We have a long road ahead before we reach, as a field, the understanding and the rigorous methodology that Harkness envisioned. This paper will try to suggest ways to help us get there.

On the Different Uses and Users of the Term Adaptation **Dorothee Behr, *GESIS – Leibniz Institute for the Social Sciences***

Transferring a questionnaire from one language and culture into another language and culture calls for translation and/or adaptation of the questionnaire. Whether translation or adaptation is required or referred to depends on various factors, among which: 1) the goal of the research (e.g., comparative), 2) the original design of the source questionnaire and, thus, its transferability to other languages and cultures (e.g., source questionnaire was designed with only one culture in mind), 3) the (linguistic) unit referred to (e.g., word vs. sentence), 4) the discipline – including its terminology – to which a researcher belongs (e.g., psychology, translation studies), or 5) personal views on what adaptation and translation involve. Firmly embracing Janet Harkness' work on adaptation (e.g., 2010), this presentation will look into the different uses and users of the term adaptation, in contrast to the term translation. This study shall encourage, in the long term, the use of a consistent terminology. A consistent understanding of what translation and adaptation involve is essential given the widespread use of cross-national research data, the different analysis techniques that can go with translation/adaptation, and the impact that different understandings of translation/adaptation have on the actual "translation" process. In the short or medium term, the aim is to raise a greater awareness of how the term adaptation, in contrast to translation, is used by different researchers. Also, a greater debate shall be encouraged on what kind of changes in translation are possible, or even required, to produce an equivalent questionnaire in comparative research. References: Harkness, Janet. (2010). VII. Adaptation of Survey Instruments. Guidelines for Best Practice in Cross-Cultural Surveys. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan.

Enhancing the Translatability of the Source Questionnaire in the European Social Survey (ESS) – Does Advance Translation Help? **Brita Dorer, *GESIS – Leibniz Institute for the Social Sciences***

To assure comparability of measurement across countries, the quality of questionnaire translation into the target languages in cross-cultural surveys is of utmost importance. Not only translation procedures and guidelines, but also the source questionnaire has an impact on the quality of the resulting translations. Starting from this, Janet Harkness developed the idea of performing ex-ante translations of a pre-final version of the source questionnaire, to be used as a 'problem-spotting tool' in order to improve the translatability of source questionnaires. The European Social Survey (ESS) was the first major social sciences survey to apply systematic 'advance translations': In its 5th and 6th round (2010 and 2011), advance translations were carried out in order to get input from people with different cultural and linguistic backgrounds. The participating teams were asked to comment primarily on translation-related problems, from linguistic or grammar issues to wording, meaning or intercultural aspects. The advance

translations led to numerous suggestions for modifications in the final English source questionnaire in both rounds. At least one third of the items advance translated were modified as a result, either by amending the wording of the source text to facilitate translation, or by adding footnotes to clarify words or expressions. This paper will evaluate the methodology applied in both rounds and describe the changes made to the final source questions following advance translations. To assess the usefulness of this method empirically, tests using Think-Aloud-Protocols (TAPs) where pre- and post-advance translation versions of questions will be translated (aloud) into German and French to evaluate whether translation of the pre- version is more problematic than translation of the post- version. Then, respondents in both target languages will be asked to think aloud while answering these questions, to evaluate potential translation effects on question processing.

Adapting Translation of the American Community Survey in Chinese and Korean **Mandy Sha, *RTI International*; Hyunjoo Park, *RTI International*; Yuling Pan, *U.S. Census Bureau***

Chinese and Korean are among the top five non-English languages for which the U.S. Census Bureau provided language assistance during the 2010 Census. However, questionnaire translation in these two languages is less studied compared to Spanish translation. This paper fills this gap by investigating unique challenges of questionnaire translation in these two languages and by providing comprehensive guidelines for translating questionnaires in Chinese and Korean. Based on results from cognitive pretesting with monolingual Chinese and Korean speakers in the United States, this paper highlights several important steps that were taken to adapt the translation to ensure functional equivalency in the translated questionnaire. This paper is based on a study that the Census Bureau undertook to translate the American Community Survey (ACS) in Chinese and Korean, in the form of a self-administered language assistance guide (LAG). Using findings from the ACS LAG study, we will discuss several issues: 1) adapting unique Chinese and Korean linguistic practices (e.g. inability to adopt common formatting stimuli in English language questionnaires such as all caps, because the Chinese and Korean writing systems are not alphabet based); 2) adopting linguistic rules (e.g. use of Hancha-rooted words and phonetic expressions in Korean); 3) adding pragmatic contextual considerations (e.g. cultural expectation when asking the marital status question); and 4) choosing appropriate translated words to reflect the immigrant experience (e.g. questions concerning migration). We will also discuss some translation difficulties that simply cannot be “fixed” within the parameters of the translation and must be addressed at the source language questionnaire level, which echoes Harkness (2003). This paper is of interest to questionnaire designers who survey non-English speakers in the United States. And our recommendations have methodological implications for translating questionnaires for other Asian languages and cultures, as well as languages that use non-Roman letters.

Translation Versus Adaptation: Translating U.S. Educational Level Survey Questions into Spanish **Patricia Goerman, *U.S. Census Bureau*; Leticia Fernández; Rosanna Quiroz, *RTI International***

Various studies have shown the difficulty of translating concepts related to country-specific programs for use in surveys. Questions about educational attainment are an example of a concept that is very difficult to translate for use with respondents with different national origins.

This is particularly the case for Spanish-speaking respondents in the United States, who come from a variety of different countries where educational systems are different not only from the U.S. system but from each other as well. This paper presents results from the cognitive testing of the Spanish translation of educational level questions in the U.S. Census Bureau's American Community Survey (ACS). Two iterative rounds of cognitive testing were conducted on a series of educational level questions with 46 Spanish-speaking respondents from 11 different countries. We found that Spanish speakers interpreted many of the educational level categories differently from what was intended. For example, Mexican-origin respondents interpreted "escuela secundaria," the original translation used for "high school," to correspond to nine years of schooling, while in the U.S. completing high school corresponds to 12 years of schooling. Similarly, while the translation for "bachelor's degree" or "bachiller universitario," was interpreted appropriately by Puerto Rican Spanish speakers, this was not the case among respondents from Argentina, Mexico, Colombia and Nicaragua. In these Latin American countries the term "bachillerato" is used to describe either junior high school or high school. Both of these translations could result in upward biases in reports of immigrant educational levels since both misinterpretations involve respondents reporting lower levels of education as higher ones. We discuss various approaches taken to deal with the comprehension differences and the extent to which these were successful. The paper concludes with a discussion of implications for translation and testing of educational levels and other country specific programs, and provides recommendations for future research.

The Origins and Development of Survey Research

The Origins and Development of Cross-National Survey Research: The Diffusion of an Innovation

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

This paper examines the rise and diffusion of survey research from the 1930s to the 1960s. It covers 1) the emergence of cross-national, survey research including the role of early adopters—Gallup, the National Opinion Research Center (NORC), other survey-research organizations, and *Public Opinion Quarterly*; 2) the initial diffusion of survey research by Gallup, International Research Associates, Inc., and others, 3) foundational survey-research meetings and associations, 4) the impact of World War II, 5) the role of the United Nations and other international organizations including its collaboration with the World Association for Public Opinion Research, 6) the first comparative surveys, 7) the contributions of international exchanges and immigrations, 8) changing developments in the 1950s and 1960s, including the role of American influence and center/periphery diffusion, and 9) impediments to development.

A History of Survey Research and Its Professional Associations

Michael Mokrzycki, *Mike Mokrzycki Survey Research Services*

The development of survey research is viewed primarily through the development on AAPOR starting with the Central City Conference in 1946.

Early Studies of Political Behavior in the United States

Michael W. Traugott, University of Michigan

Political polling has been central to the development of survey research and promotion of its adoption in the United States and in other countries. This presentation focuses on the distinctive roles of pollsters, academics, and news organizations and their involvement with political polling as a critical element in the development of survey research in the United States. This is a story of institutional conflicts and research design differences, and the ways they affected the advancement of knowledge about polling methodology as well our understanding of political behavior. It also explains a series of paradigmatic shifts in models explaining how and why people vote. Across 75 years of development, relations between academic and commercial pollsters have waxed and waned. In contemporary polling, academics continue to provide most of the methodological development, quickly adopted by commercial pollsters.

A History of Survey Research at NORC

Norman Bradburn, Department of Psychology, NORC at the University of Chicago;
James A. Davis, Department of Psychology, NORC at the University of Chicago

The National Opinion Research Center (now just NORC) was founded by the Harry Field at the University of Denver in 1941. Breaking from the commercial orientation of industry founders Gallup, Roper, and Crossley, the National Opinion Research Center aimed to do survey research in the public domain and to serve the social science community. Field in 1946 organized the first survey research conference at Central Center which led to the founding of AAPOR.

Comparing Early Survey Research Methodologies in Mexico in the 1940s

Alejandro Moreno, Instituto Tecnológico Autónomo de México

In this paper I compare the development of public opinion research in Mexico during World War II in different areas: media polls, academic research, and policy-oriented surveys. The latter two include the works by Laszlo Radvanyi at the National University's Scientific Institute of Public Opinion, as well as various works sponsored and conducted by the U.S. State Department. Early polls developed by media outlets both in Mexico City and in Monterrey illustrate a continuous measurement of public opinion, with ad-hoc methodologies that were still far from proper probability sampling and questionnaire design, but that gave voice to politically excluded segments of the population, such as women, who were granted the right to vote until 1953. For the academic research efforts, I analyze some of the contents published in the *International Journal of Attitude and Opinion Research*, edited by professor Radvanyi in Mexico City and published for the first time in 1947. The word "encuesta," used indifferently in Spanish for polls and surveys, was used in the 1940s in academic books that relied on interviews with experts rather than a broader public, but referring to a process of interviewing "some" sample. In addition to the methodologies, this paper pays a special attention to the social groups that were not only used in the stratification of samples but also in the analyses of poll results, providing ways in which Mexican society was conceived during those years.

Maximizing Response Through Optimal Contact Strategies

Number of Mail and Phone Contact Attempts to Complete Physician Surveys

Julie C. Linville, *SRA International*; Eric Jamoom, *National Center for Health Statistics*; Paul C. Beatty, *National Center for Health Statistics*; Nicholas A. Holt, *SRA International*

The National Center for Health Statistics has conducted the Electronic Health Records Supplement (EHRS) of the National Ambulatory Medical Care Survey (NAMCS) annually by mail since 2008. The EHRS asks physicians about their use of electronic health records (EHRs), with 10,302 physicians surveyed each year in 2010, 2011 and 2012. A sample of 5,232 respondents to the 2011 EHRS were impaneled for the subsequent Physician Workflow Study (PWS), a three-year longitudinal study designed to obtain additional information from physicians regarding the impacts and barriers to adopting EHR systems. The PWS was administered using a modification of methods proposed by Dillman (2007), with up to three mailings of the questionnaire, a reminder postcard sent after the first mailing, and telephone follow-up of non-respondents. Two versions of the PWS were developed, one for EHR adopters and another for EHR non-adopters. Although the sampled physician was the intended respondent, proxy responses from staff members in the practice were accepted when necessary. This paper will examine the relationship between contact history and response for three years of the EHRS (2010-2012) and the first two years of the longitudinal PWS (2011-2012). We will explore the response yield and efficiency from each wave of contact. We will also analyze differences across surveys, across adopter and non-adopter strata, across physician specialties and across other practice characteristics. In addition, we will investigate the relationship between contact history, physician specialty and adopter status on the prevalence of proxy response. Finally, we will consider the implications of these findings toward the most efficient approaches for maximizing responses from the sampled physicians.

Issues in Contacting and Engaging SNAP Recipients in a Longitudinal Survey

Crystal MacAllum, *Westat*; Suzanne McNutt, *Westat*; Adam Chu, *Westat*; Susan Bartlett, *Abt Associates*; Kelly Kinnison, *USDA Food and Nutrition Service*

The Supplemental Nutrition Assistance Program (SNAP) provides nutritional foods to low-income families. The Food and Nutrition Service (FNS) in the U.S. Department of Agriculture administers the program. The FNS Healthy Incentives Pilot (HIP) Evaluation assessed the impact of giving a financial incentive for the purchase of fruits and vegetables on recipients' diet. A random sample of 2,538 SNAP recipients in one U.S. county received the incentive while a comparable sample of 2,538 in the same county did not. The study attempted to contact and engage sampled SNAP recipients in three rounds of interviews over a 16-month period, spaced approximately every six months. SNAP recipients are a mobile population that is hard to reach and engage in research; therefore, obtaining and retaining a sample large enough to achieve adequate power to detect a change in diet at each round of interviews was a challenge for the study. This paper presents the strategies employed to contact and engage this population, including telephone data collection with in-person field follow-up for non-respondents and those with missing telephone numbers; progressively greater incentives over rounds of data collection; additional incentives if respondents used their cell phones; and the use of iPads to manage the interplay between cases in the field and those in the telephone center. The study was successful in increasing the proportion of telephone respondents over waves: In Round 1 58% of interviews were field completes and 42% telephone completes; by Round 3, only 26% were field completes while 74% were telephone completes.

Improving Response and Operational Efficiency Under the Constraints of Time-Sensitive Program Evaluation

Andy Weiss, *Abt SRBI*; Rhoda Cohen, *Mathematica Policy Research*; Faith Lewis, *Abt SRBI*

Surveys to evaluate government benefit programs are constrained by the administrative structure of those programs. Issues like time-sensitive administration and access to program participants limit survey design choices. This can be an especially complicated problem for locally administered programs. Flexibility in adapting survey design to local conditions holds promise for improving response rates and other quality measures. We conducted an evaluation of the USDA's Summer Electronic Benefit Transfer for Children (SEBTC) program, which provides a supplemental nutrition benefit to households with school-aged children during the summer. In 2011, the sample included 5 sites and interviews with 5,000 households before the school year ended and again in the summer. In the 2nd year, the evaluation entailed collecting data from 14 sites and interviews with 27,000 households during each wave. All interviews were completed over the telephone. Respondents were contacted by mail distribution of a toll-free call-in number, outbound phone calls and in-person interviewers going to respondent homes and initiating a call to our call center. The random assignment study assessed the impact of SEBTC on children's food security and other nutrition related measures. By implementing a wide range of methods from providing technical assistance to help school districts consent households to a case-level customized calling algorithm we improved the response rates from 65% in the summer of 2011 to 80% in the summer of 2012.

Setting Expectations for Managing Interviewer Performance

Barbara C. O'Hare, *U.S. Census Bureau*; Tamara S. Adams, *U.S. Census Bureau*; Chandra Erdman, *U.S. Census Bureau*; James B. Lawrence, *U.S. Census Bureau*

Differential survey response across subpopulations and geographic areas is well documented in the survey literature. The current challenges of survey administration require setting realistic expectations of survey response rates, particularly in assessing survey progress and deciding where to direct data collection effort. This paper discusses the development and implementation of standardized field interviewer performance standards based on statistical variation in demographic and socio-economic characteristics of neighborhood (block group) areas. The result of this effort is a set of standards driven by the neighborhood characteristics of the interviewer's cases, rather than the overall county response rate of where the interviewer primarily works. We defined the new field CAPI interviewer response rate standards through statistical analysis of census and American Community Survey block group data to identify the best predictors of census and survey response, to cluster neighborhoods, and to determine the optimal number of performance strata. (Note: the analysis is described in detail in an abstract submitted by Erdman, Adams, and Lawrence). After establishing new strata that reflect variations in response rate, we addressed the practical issues of implementing new standards on which interviewers are evaluated, including: 1) The process of establishing a distribution of expected response rates within each stratum used to define five performance levels, adjusting for small caseloads. 2) The collaborative effort between the field operations staff and the statistical analysts to set the boundaries of the performance levels. 3) The challenges of presenting the statistical analyses to field operations staff and addressing the practical concerns of the need for face validity and clarity of how the standards were set to address interviewer and

personnel policy concerns. The experiences presented here can be of value to other survey organizations setting survey performance expectations, in that it highlights the challenges of the practical operational challenges of implementing statistically derived standards.

First Contact Strategies for Web Surveys: Is a Phone Call or a Letter the More Effective Introduction?

Jill Connelly, *NORC at the University of Chicago*; Micah Sjoblom, *NORC at the University of Chicago*; A. Rupa Datta, *NORC at the University of Chicago*; Peter Hepburn, *NORC at the University of Chicago*

The objective of the National Survey of Early Care and Education (NSECE) is to document the nation's current use and availability of early care and education, and to deepen our understanding of the extent to which families' needs and preferences coordinate well with providers' offerings and constraints. The NSECE included a survey of home-based child care providers who were licensed or otherwise registered with state agencies. The survey included Web data collection, with phone or in-person follow up as needed. Individuals who provide care to children in a home-based setting tend to be older or lower-income or in other demographic subgroups that have lower Internet usage rates. In order to encourage participation by Web, a \$35 gift card was offered to complete the interview online. We had phone numbers, but no mailing or email addresses for sampled individuals. We designed an experiment with 1,300 providers to test whether it would be more efficient to 1) send a letter or email as a first contact based on locating efforts that didn't involve personal contact with the respondent, or 2) make a gaining cooperation phone call first, to introduce the study and then request mailing or email information to send the Web survey request. Our evaluation includes comparisons of effort required, success rates in reaching respondents through initial contact attempts, cooperation with the initial request, and final cooperation rates.

Incentives and Survey Response

Survey Incentive Fees, Data Quality, Nonresponse, and Survey Administration
Jesse Bricker, *Federal Reserve Board of Governors*

This paper uses both the 2007-2009 Survey of Consumer Finances (SCF) panel and the 2007 and 2010 SCF cross sections to investigate whether monetary incentives help data quality, conditional on responding to the survey, and help reduce time in the field. The 2007 SCF had a base incentive of \$20, though many needed \$50 before responding; the base incentive in 2009 was \$50. This first component of this paper compares the response rates and data quality of two groups: those that received a \$50 incentive in both waves to those that received a base \$20 incentive in the first wave and the \$50 base incentive in the second. Data quality is measured by the use verifying documents and the precision of responses. The second component of this paper uses two levels of variation to investigate the impact of incentives on time that field staff spends in the field and the total number of times that field staff contacted potential respondents. First, we use variation in base incentive across the 2007 and 2010 SCF cross sections, as the base 2010 incentive was also increased to \$50. Second, there is also variation across sampling regions, as cost of living varies across rural and urban areas. Conditioning on detailed local information and focusing on sampling areas that were used in both the 2007 and the 2010 surveys will allow us to analyze the costs and benefits of a larger survey incentive on time spent in the field collecting data.

Timing of Nonparticipation in an Online Panel: The Effect of Incentive Strategies
Salima Douhou, CentERdata, Tilburg University; Annette Scherpenzeel, CentERdata, Tilburg University

Nonresponse in (online) panel surveys is problematic since it may lead to a bias. An important measure to secure respondent cooperation is the use of monetary incentives. An experiment was carried out in the LISS panel (Longitudinal Internet Studies for the Social Sciences, an online panel based on a true probability sample of households) in 2007 to determine the optimal recruitment strategy for a new online household panel (see Scherpenzeel and Toepoel, 2012). The monetary incentives varied during the recruitment. The incentives were either promised or prepaid and the amount varied (10, 20 or 50 euros). More than 500 respondents were randomly selected in the different incentive conditions. The prepaid incentives were quite effective in increasing the recruitment rates for the panel. However, a question often posed is how these incentives affect the long term participation in the panel. Are the respondents who were recruited with the help of the high incentives not dropping out faster than respondents who participate with intrinsic motivation? The purpose of this paper is to find out which incentive strategy is efficient for long term participation of respondents, five years after the recruitment. Efficiency implies both low recruitment costs combined with high response rate after entrance in the panel. This paper takes a different approach to model the time-to-event of nonparticipation: survival analysis. The event in this case is nonparticipation. This method has two important advantages: 1) incorporates the timing of the event and 2) allows for censoring. This research will provide new evidence on the timing of nonparticipation and the influence of different incentive strategies on this timing. The paper will present the willingness of respondents to participate for a long term in the panel for different incentive strategies.

Nonresponse and Nonresponse Bias in a Probability-Based Internet Panel: The Effect of (Un)conditional Cash Incentives
Annelies Blom, University of Mannheim; Ulrich Krieger, University of Mannheim

The German Internet Panel (GIP) is a new large-scale online panel based on a probability sample of individuals living within households in Germany. In 2012 households were approached offline, with a short face-to-face interview. Subsequently, all household members were invited to complete the bi-monthly GIP questionnaires. To minimize non-coverage bias, households without access to the Internet were provided with the necessary hardware and/or a broadband Internet connection. Recruitment into the GIP consisted of various stages: the face-to-face household interview, mailed invitations to the online survey, reminder letters, a phone follow-up, and final mailed reminders. During the face-to-face phase we conducted an experiment with €5 unconditional vs. €10 conditional household incentives. In addition, an experiment with €5 unconditional personal incentives was conducted during the first reminder. We examine the question of whether a carefully recruited, probability-based online panel can be representative of the general population and is thus suitable for social and economic research. The models presented analyze the processes leading to participation and associated biases in the sample. The various stages of recruitment into the GIP are assessed, as well as the effects of the two incentives experiments.

The Effect of Prepaid Incentives on Responses to Sensitive Questions in a Mail Survey

Rebecca Medway, *American Institutes for Research*

Researchers have expressed concern that offering monetary incentives in surveys may have unintended effects on responses to survey questions. The current literature exploring the effect of incentives on response distributions finds limited support for this fear. However, when researchers have investigated the impact of incentives on survey responses, they typically have analyzed all of the survey items as one group. It is possible that the incentive effect varies depending on item characteristics, and that the decision to analyze all of the items at once masks significant differences for subgroups of items. In particular, responses to sensitive items appear to be subject to situational factors and survey design features; as a result, these items may be more susceptible than non-sensitive ones to incentive effects. Furthermore, research repeatedly shows that respondents misreport for sensitive items, so it would be useful to know whether incentives affect how honestly respondents answer such items. This paper explores the effect that offering a prepaid cash incentive had on self-presentation concerns and responses to sensitive questions in a mail survey of registered voters. As compared to a control group that did not receive an incentive, respondents who received \$5 reported a significantly greater number of highly sensitive, undesirable attitudes and behaviors. The incentive had no effect on responses to less sensitive items, suggesting that item sensitivity may play a role in the magnitude of the incentive effect. For three voting items where validation data was available, the incentive resulted in a general pattern of reduced nonresponse bias and increased measurement bias; however these effects generally were not significant. The effect of the incentive did not vary significantly by respondent characteristics.

Effective e-incentive for Online Study: Comparing Branded e-Gift Card and Virtual Cash Card

Teresa (Ye) Jin, *The Nielsen Company*; Shu Duan, *The Nielsen Company*; Jennie Lai, *The Nielsen Company*; Michael W. Link, *The Nielsen Company*

Given the continued growth of young adults with Internet access, the incentive method should complement the survey mode especially online studies for repeated measures. Past empirical research examined potential online incentive methods such as vouchers, lotteries or donations, eGift card, or virtual incentive and studied their effect on response rates. In an effort to gain greater cooperation among the hard to reach cohorts (e.g., in particular young adults), Nielsen will administer an online study using a Web-based application to collect media consumption behavior. The research objectives are two-fold: 1) to test the effectiveness of incentive methods (choice vs. no-choice) and 2) e-incentive options (branded e-gift card incentive vs. virtual cash card incentive). The address-based sample will be randomly assigned to three conditions: branded eGift card (i.e., Amazon.com Gift Card), virtual cash Visa card, or respondent can choose any one of the two options listed above. The qualified respondent in the household will be asked to participate in the one-week online study. This research paper will evaluate cooperation rate by key demographic characteristics and compliance during the data collection period for each incentive condition. The research findings will advance the body of knowledge on the most effective incentive method and option to gain cooperation of the hard-to-reach cohort in the digital age of online usage.

Friday, May 17
1:45 p.m. – 3:15 p.m.
AAPOR Concurrent Session E

Developments in the Design and Implementation of Web Surveys

The Effect of Compressing Questionnaire Length on Data Quality

Jessica LeBlanc, *Center for Survey Research at University of Massachusetts Boston*;
Carol Cosenza, *Center for Survey Research at University of Massachusetts Boston*

Consumer Assessment of Healthcare Providers and Systems (CAHPS®) instrument guidelines recommend formatting ordinal response categories vertically. However, in an effort to create questionnaires with fewer pages, some users have formatted the response options horizontally. Frequently, when there are too many answer categories to fit on one horizontal line, CAHPS users format responses horizontally over multiple rows. This formatting may lead respondents to search for an appropriate response in the scale. As part of a survey of adult patients from a university-based health system (n=2100), a methodological experiment was implemented, with respondents randomized to receive one of three versions of the questionnaire. Version A maintained closer compliance to CAHPS guidelines, containing mostly vertical scales with horizontal scales used only when fit onto a single line. Versions B and C both contained response scales with multiple columns and rows. In version B, ordinal response options were listed horizontally in two rows (read from left to right, top-bottom) and in version C, response options were listed vertically in two columns (read from top to bottom, left-right). Analysis of this data will focus on differences among versions A, B, and C in survey response rates, mean scores for single items, and item non-response. Particular attention will be paid to differences between items that ask respondents to report on frequency of events (e.g. number of doctor visits) or demographics (e.g., age) and items that ask respondents to assess experiences using adjectival scales (e.g., never, sometimes, usually, always), which may be more difficult for respondents to choose when the presentation of the ordinal responses is disrupted. This test used the CAHPS® Clinician & Group Patient-Centered Medical Home adult questionnaire and utilized a standard 3-contact mailing protocol. It was funded by the Agency for Healthcare Research and Quality. Data collection was completed in 2011.

Evaluating Interactive Feedback in Computer-Assisted Self-Interviewing (CASI)

Margaret L. Hudson, *University of Michigan*; **Andrew L. Hupp, *University of Michigan*;**
Chan Zhang, *University of Michigan*; **Heather M. Schroeder, *University of Michigan***

A long-standing concern with self-interviewing methods is that respondents may lack the motivation to spend effort in completing the survey, which can lead to satisficing and compromised data quality. Recently researchers have started to explore the use of interactive feedback in computer-assisted self-interviewing (CASI) whereby respondents are prompted if satisficing behaviors are detected (e.g., respondents receive messages saying they are going too fast when their response time is quicker than a certain threshold). In particular, a small number of studies, mostly using online panels, have shown that such interactive feedback can effectively reduce targeted undesirable behaviors in Web surveys without a substantial increase in break-offs. While these findings are promising, it is not clear if the same success would be

observed with other survey populations who may not be as motivated to complete surveys as panel respondents. Even more importantly, little is known as to whether this type of interactive feedback in self-administered surveys could affect perceived privacy and thus, introduce social desirability bias in answers to sensitive questions. We will report findings from a CASI survey of mental health risk and resilience among Soldiers new to the U.S. Army. Response speed prompts were implemented in response to concerns about satisficing behavior. The speed prompts were introduced approximately one quarter of the way through the study. Since the monthly samples are independent and representative, a natural pre/post comparison is possible. Survey data will be compared before and after the implementation to evaluate whether these prompts can effectively influence response time and improve response quality (based on indicators such as item nonresponse, straightlining, and acquiescence). We will also assess if the use of prompts could backfire – i.e., producing more break-offs and fewer reports of socially undesirable answers, given the survey is voluntary and contains many sensitive questions (e.g., suicidal ideation).

Are You Seeing What I am Seeing? Exploring Response Option Visual Design Effects With Eye-Tracking

Amanda Libman, *University of Nebraska – Lincoln*; Jolene D. Smyth, *University of Nebraska – Lincoln*; Kristen Olson, *University of Nebraska – Lincoln*

Since the late 1990s theory drawn from the vision sciences and Gestalt psychology has guided the visual design of questionnaires. A considerable amount of research has been conducted that shows that altering questionnaire visual design can change response distributions and data quality (Dillman, Smyth, and Christian 2009; Jenkins and Dillman 1995; Tourangeau, Couper, & Conrad 2004). However, this research is limited in what it can tell us about how different visual designs influence responses. In other words, the evidence for how visual design matters is largely circumstantial. Eye-tracking technology gives us the opportunity to overcome this challenge. A handful of studies have used eye tracking to better understand how respondents see and process a questionnaire (Galesic et al. 2008; Lenzer, Kaczmirek, and Galesic 2011). In this paper, we will explore how visual design in response options assists respondents in processing survey questions. Specifically, we will analyze eye-tracking data to examine the effects of Web survey response option experiments that include symbolic language, grid response options and the use of single and double columns. Preliminary evidence from the lab shows that the addition of smiley faces to a Likert scale cause respondents to slow down when processing the given response options. By observing how respondents actually view the different versions of the questionnaire and visual aids, this study will contribute to our understanding of how and why visual design influences responses and will shed light on best practices for questionnaire design.

Classifying Mouse Movements to Predict Respondent Difficulty

Rachel Horwitz, *U.S. Census Bureau*

A goal of the survey interview is to collect reliable and valid data. Achieving this goal is often difficult because respondents may not understand what is being asked of them. In traditional interviewer-administered survey modes, interviewers can pick up on signs of confusion and difficulty answering a question from the respondent's speech patterns, expressions, or response times. In self-administered surveys, however, identifying confused respondents has previously not been possible. The introduction of Web surveys provides an interactive environment with a

vast amount of data that researchers can collect in real time. Using these data, it may be possible to determine when respondents are having difficulty answering a question, much like in an interviewer-administered survey. Using Web browsing and education research as a basis, this paper identifies 11 unique movements that respondents make with the mouse cursor while answering survey questions. Through an exploratory analysis, we hypothesized which of these movements are related to difficulty answering survey questions. Then, using scenarios to manipulate question difficulty and asking participants to rate the difficulty of each question, we were able to test our hypotheses to determine which movements are related to difficulty and which are general movements people make when interacting with a computer. Finally, this paper proposes a model that can be used to predict, in real time, when a respondent is having difficulty answering a survey question. We find that not only are certain mouse movements highly predictive of difficulty, but they are more predictive than response times, which have been used to predict difficulty in the past. This information can be used to provide real-time help to confused participants or it can act as a diagnostic tool to identify confusing questions.

Dynamic Visual Design for List-Style Open-Ended Questions in Web Surveys **Marek Fuchs, *Damstadt University of Technology***

Several studies have demonstrated that respondents react to the size and design of the answer field offered with open-ended questions in Web surveys. Larger answer boxes seem to pose an additional burden and yield fewer answers and higher rates of item nonresponse as compared to smaller answer boxes. At the same time larger answer boxes work as a stimulus that increases the length of the response provided by those respondents who actually answer the question. Similar findings have been demonstrated for list-style open-ended questions where respondents are supposed to type short responses (e.g. name of countries, cities, or brand names). In this paper we evaluate a method optimizing the extent of the answer to list-style open-ended questions without increasing item nonresponse. We use a dynamic screen design where respondents initially were exposed to one fixed answer box. If respondents entered a response into an initially visible answer box, a second answer box appeared. If they again entered a response a third box appeared (and so on). In a randomized field-experimental study embedded in a large scale survey (n=6,100) we tested several question versions combining various numbers of fixed and dynamic answer boxes in a between-subjects design. Results indicated that the optimal design consisted of three initially visible (fixed) answer boxes and dynamically providing further answer boxes if the respondents wanted to answer more extent. Findings are discussed in light of the impact of the dynamic visual design on the question answer process.

Question Order and Context Effects

Question Order Effects on Estimates of the Size and Characteristics of Religious Groups

**Gregory A. Smith, *Pew Research Center*; Besheer Mohamed, *Pew Research Center*;
Jessica Hamar Martinez, *Pew Research Center***

Religious identification is a key variable for understanding opinion on many topics (including politics and elections) and a multifaceted, complex concept. It is an indicator of the religious groups with which one identifies and one's religious beliefs. But it can also tap ethnic or cultural attachments even in the absence of any ongoing religious commitment. Depending on their

goals, researchers may be interested in one or another of these aspects of religious identification. Some will only be interested in those who currently identify religiously with a group. Others will be interested in the broader group of those who identify in some way with a group (e.g., by virtue of their upbringing or family background) even if they do not currently think of themselves as Catholic or Jewish or Mormon (for example) in religious terms. We report the results of experiments in which we varied the wording and order of questions about religious affiliation. We show that the wording and order of religious affiliation questions can have a substantial impact on estimates of the size of religious groups. And we assess the degree to which varying approaches to question wording and order can produce different estimates of the religious attributes and demographic characteristics of religious groups. We discuss the ways in which the specifics of how one asks about religion can shape the results one obtains. The findings of our study are consequential not just for scholars of American religion, but also for the many social and political researchers who include religious identification in their surveys or as a variable in their studies.

Context Effects in Candidate Favorability Ratings: Lessons From the 2012 Elections

Eran Ben-Porath, *Social Science Research Solutions*; Damla Ergun, *Langer Research Associates*; Gregory Holyk, *Langer Research Associates*; Gary Langer, *Langer Research Associates*; Jon Cohen, *Capital Insight/Washington Post Media*

This study builds on context effects theory to test the impact of question order during an ongoing favorability measurement of presidential candidates. Throughout the primaries and the campaign leading up to the 2012 presidential election, respondents were asked how favorable they felt toward various candidates. Our findings indicate that when respondents were asked about Mitt Romney after Barack Obama, Romney was consistently rated more unfavorably than when his name came before Obama's. When respondents were asked about Romney prior to Obama, the share of "Don't Know" responses to Romney was significantly higher as was his relative popularity among those with an expressed opinion. Order-effects for Obama's favorability ratings were not apparent. A similar pattern occurred in other comparisons as well. For example, when asked about Rick Santorum prior to Mitt Romney, there were significantly more "Don't Know" responses for Santorum than when respondents were asked first about Romney. Both favorable and unfavorable responses to Santorum were higher when Romney was the first candidate mentioned. These findings illustrate how asking about the more familiar candidates first provides context for the lesser-known candidates. Consistent with previous research on context effects, order effects were strongest among respondents with less education, lending further support to the idea that the better-known candidate provides a context in which assessments are made. The results suggest that question order variations could result in measurement of two different types of attitudes about lesser-known candidates: one is the "true" attitude in so far as it can be measured, while the other is the attitude relative to the better-known candidate. This, in turn, raises questions as to whether one order or another better approximates the actual construct of favorability, and when rotating question order is appropriate, given specific research objectives.

Interaction Between Question Context Effects and Linguistic Backgrounds **Sunghye Lee, *University of Michigan*; Norbert Schwarz, *University of Michigan***

Despite lacking theories, question context effects are one of the most frequently examined measurement errors. Based on social cognition and communication theories and the notion of high vs. low context culture, we hypothesized 1) interactions among textual, cultural, and external question contexts. We chose the self-rated health (SRH) question, a popular survey item believed to be immune to context effects, and further hypothesized 2) larger context effects for Spanish speakers (and Hispanics) than English speakers (and non-Hispanics). We conducted two sets of experiments in a multilingual survey. A subset of respondents was randomly assigned to different textual contexts of SRH by varying its order in a questionnaire. The results supported the hypotheses. English-speaking respondents' reports on SRH were consistent across all textual contexts, but simple changes in the textual contexts produced dramatically different reports by Spanish-speaking respondents. Specifically, Spanish speakers reported substantially better health when SRH was asked after specific health condition questions than before any health-related questions. Because language is a proxy for culture, this demonstrated an interaction between textual and cultural contexts. Furthermore, among Spanish speakers, the textual context effects were larger for females and older respondents and differed by comorbidity status, illustrating an interaction among three types of contexts. Implications are twofold. First, context effect patterns observed in one culture do not necessarily apply to another culture. Second, even within the same culture, context effects vary by respondents' characteristics. Hence, context effects studied with a homogeneous group should not be assumed to hold in cross-cultural studies.

Some Informal Experiments on the Effects of Questionnaire Design Changes on Item Nonresponse

Christine Kudisch, *Experian Marketing Services*; Josephine Leonard, *Experian Marketing Services*; Max Kilger, *Experian Marketing Services*; Charlie Palit, *University of Wisconsin-Madison*

For decades, Experian Simmons has conducted a national survey of U.S. consumers reporting data on a sample of approximately 25,000 adults age 18+ annually. The mail survey instrument is particularly large in scope and widely varied in topical content as well as broadly diverse in the types of question formats used to measure those topics. This has provided us with opportunities to empirically investigate the effect of different ways of asking questions and their impact on item nonresponse. As a result we have collected an interesting set of examples illustrating how changes in question wording and position can affect item non-response. This presentation will present and discuss some common features of specific question formats that affect item non-response. We also present the results of informal experiments aimed at reducing item non-response bias through modifications to the question format as well as question positioning changes. In addition we examine aspects of differential non-response to questions by ethnicity in an exploration of some specific question formats that differ in terms of non-response by Hispanic, non-Hispanic status and by language preference among Hispanics.

Are Question Context Effects Partially A Function of Forced Choice Questions? **David Moore, *University of New Hampshire***

Crucial to a sustainable future for public policy polls is whether they provide meaningful assessments of public opinion. Often polls on even the same subject, however, produce contradictory results, which are explained by attributing the differences to question wording or question context effects. This paper reports on two different representative surveys that show 1) a particular response order effect and, separately, 2) a particular question wording effect, were not present among people with “intense opinions,” though the two effects were found for the overall samples. These results suggest that if polls were measuring “meaningful” (i.e., intensely held) opinions, some (or many) of the contradictory results produced by polls would disappear. Background: Many polls ask public policy questions that pressure respondents to produce an opinion, even if they don’t have one. The result: Typically more than 9 in 10 Americans appear to have a meaningful opinion about virtually all issues. Separately, polls on the same subject often produce startlingly different results. In May 2011, five polling organizations all asked about bringing home troops from Afghanistan in the wake of Osama bin Laden’s death. Two polls showed strong majorities in favor, two showed about an evenly divided public, and one found strong opposition. A frequent explanation for such contradictory findings is that small differences in question wording and question order could produce major differences in results. But maybe that’s at least partly because we include people who really don’t have opinions, but are pressured to respond anyway, and who are therefore particularly susceptible to small differences in question wording and question order. The experiments reported in this paper suggest that notion has merit. Both the response order effect and question wording effect were minimized (or eliminated) when only people with intense opinions were analyzed.

Multi-cultural and Multi-Lingual Survey Research

A Comparison of Hispanic Households That Were Identified by Hispanic Surname to Those That Were Not

Dan Estersohn, *Arbitron Inc.*; Kelly Dixon, *Arbitron Inc.*; Mike Kwanisai, *Arbitron Inc.*; Al Tupek, *Arbitron Inc.*

In partnership with our sampling vendor (SSI) Arbitron has been investigating potential uses for sample that has been appended with demographic data. The most useful attribute discovered so far has been Hispanic household identification. SSI’s identification is based upon matching a householder’s last name to a Hispanic surname list. SSI’s list is based the Census Bureau’s Hispanic surnames list which has been in use (with modifications) for over 50 years. The surname matching is not a perfect identification method. Some households are incorrectly tagged as Hispanic while other Hispanic households are not identified. We propose to investigate whether the correctly tagged Hispanic households are demographically or geographically different than the Hispanic households that were not identified as Hispanic. Differences between the two groups might suggest differential contact strategies such as materials, incentives, or interviewer language. Arbitron’s respondent procedures are used to identify the actual Hispanic households. Among the Arbitron-collected variables for the comparison will be age, household size, number of persons, presence of children, the presence of non-Hispanic persons in each group of households, and language spoken most often at home. A test for spatial clustering of each group will also be performed. If one or both of the groups are spatially clustered then an analysis of neighborhood Census variables can also be undertaken. Among the neighborhood-level (census tracts) Census/ACS variables that can be

used are the Hispanic percent of the population, native vs. foreign-born, “linguistic Isolation,” educational attainment and median income.

Survey Error and Survey Costs of Interviews Using Real-Time Interpreters **Stephen Immerwahr, *New York City Department of Health and Mental Hygiene*; Tara Merry, *Abt SRBI***

Real-time interpreter services can be used to include linguistically isolated respondents in telephone surveys, but the inherently unstandardized nature of these interviews raises serious concerns about measurement error. Despite calls for evaluation, published analyses of survey error and costs associated with real-time interpretation are rare. (Hu et al. 2010 and Link et al. 2009 are two recent articles of note.) Using data from a computer-assisted telephone interviewing survey conducted in New York City between September 2008 and February 2009, we compare survey error and cost for 82 interviews conducted in multiple languages by interviewers aided by a commercial telephone interpreter service with 7472 standardized interviews conducted in English or by bilingual interviewers in Spanish, Russian, and two Chinese dialects. We report mean item nonresponse, average relative variance across continuous variables, nonresponse to specific health conditions and behaviors, and within-survey break-offs (starting but not completing the interview). We compare direct costs (translating the survey into Spanish, Russian, Chinese scripts) and operation costs (calling and interviewing, line costs, and live interpreter service fees). Overall differences in item nonresponse were small but were substantial for some individual survey measures. For example, when asked for the number of opposite-sex sexual partners in the past 12 months, 18% of real-time interpreter interviews resulted in ‘don’t know’ or ‘refused’ responses compared to 6% overall. Within-survey break-offs were also higher (25% vs. 11% overall). The cost-per-complete for real-time interpreter interviews was \$470: more than nine times the cost of interviews in English or Spanish, and roughly four times that of Russian or Chinese language interviews. The challenge facing survey researchers using real-time interpreters is to balance reduction of bias by including this population with potential measurement error and greater cost.

Resolving Multilingual Issues in Survey Development: Experiences From a Translation Workshop

Stephanie Beauvais, *Westat*; Jocelyn Newsome, *Westat*; Martha Stapleton, *Westat*; Kerry Levin, *Westat*; Salma Shariff-Marco, *Cancer Prevention Institute of California*; Nancy Breen, *National Cancer Institute*; Gordon Willis, *National Cancer Institute*

As surveys are increasingly administered in multiple languages, researchers must consider both language and culture during translation (Harkness et al. 2010). In an innovative approach to survey translation, Westat and NCI recently held a workshop that tackled multi-lingual issues across multiple languages simultaneously. Its purpose was to address previously identified problems with the Spanish- and Asian-language (Mandarin, Cantonese, Vietnamese, and Korean) versions of the California Health Interview Survey (CHIS) Discrimination Module (DM) (Shariff-Marco et al. 2009). Earlier behavior coding efforts identified a dissonance between the translations and the original intent of the English-language items (Levin et al., 2010). These translation “mismatches” were the focus of the workshop. Since both cultural and linguistic issues were to be addressed in the workshop, the project team sought “culture brokers,” rather than translators, for each language. Culture broker is an anthropological term referring to someone who mediates and facilitates understanding between cultures (Jezewski & Sotnik,

2001). Each language team was comprised of individuals who had experience with survey research, were knowledgeable about the culture and language of the target group, and were able to think critically and collaborate with others. The translation workshop was designed to focus on conceptual equivalence rather than exact word-for-word translation. The findings from the workshop identified four primary areas where translations were problematic. First, it was often difficult to capture nuances of an English idiom in translation. Second, there were instances when simply no lexical counterpart existed in other languages. Third, response scales suffered in translation. Finally, certain survey conventions posed unexpected problems. In this paper, we discuss our experiences developing the translation workshop and finding the culture brokers. We also discuss workshop dynamics and the team's resolutions for the problematic translations. We conclude by proposing areas for future research in multicultural and multilingual survey development.

Are Latin Americans as Courteous as People Say? Survey Experiment Evidence on "Courtesy Bias" From Five Countries

David Crow, *Centro de Investigacion y Docencia Economicas (CIDE)*; Gerardo Maldonado, *Centro de Investigacion y Docencia Economicas (CIDE)*

pronounced in Latin America for two reasons. First, given the relative recency of survey research in Latin America, potential respondents may be more willing to participate in surveys and more civil when they do. Second, given the formality and hospitality that characterizes interpersonal communication in Latin America, respondents may be reluctant to give unvarnished answers, preferring to put matters in the best light possible. Does courtesy bias exist in Latin America? We attempt to answer this question by means of survey experiments conducted in five Latin American countries (Brazil, Colombia, Ecuador, Mexico, and Uruguay). If Latin Americans are susceptible to courtesy bias, we should observe it in all countries—though, given intraregional cultural heterogeneity, to varying degrees. The survey experiment consists of splitting national samples and providing each half-sample with a different response set for each of four batteries of questions (on institutional trust, government performance evaluations, attitudes on immigration policy, and support for liberal intraregional trade and migration policy). The first response set was a four-point scale common in cross-national research ('a lot,' 'somewhat,' 'a little' and 'not at all') and the second, a seven-point response scale. The seven-point scale's neutral midpoint and more graduated response options, in theory, give respondents greater opportunity to nuance their ratings. We expect that courtesy bias will result in higher means on the four-point scale than on the seven-point scale (after rescaling both from 0 to 1). Statistically indistinguishable means imply the absence of courtesy bias.

Respondent Difficulty in Cognitive Interviews: From Findings of Chinese and Korean Cognitive Interviews

Hyunjoo Park, *RTI International*; Mandy Sha, *RTI International*; Murrey Olmsted, *RTI International*

Cognitive interviewing has been widely used as a tool for pretesting and improving questionnaires. As noted by Willis (2005), respondent recruitment determines the feasibility of a cognitive interview study and the selected sample should attempt to cover a cross-section of the population that is being studied. However, the practical side of conducting cognitive interviews has received little empirical attention—in particular, not much is known about how to optimize respondent selection. Literature has shown that level of education and age affect one's

cognitive ability, including recall and verbal fluency; these are all skills required for being a “good” cognitive interview respondent. Following this logic, the information produced from each cognitive interview and its utility may vary. Using interview data and associated paradata (e.g., interview length) from 258 non-English language (Chinese and Korean) cognitive interviews from the American Community Survey (ACS), this paper identifies indicators of respondent difficulty and examines how those indicators are related to the outcome of cognitive interviews. First, we identify variables likely to represent respondent difficulty and establish a profile of participants who experience higher response difficulty with interview probes and the cognitive interview setting based on the interviewers’ rating. We hypothesize that compared to those with lower response difficulty, respondents with higher response difficulty will identify a larger number of issues in cognitive interviews along with different types of issues (i.e., issues indicating valid questionnaire problems vs. issues prone to user errors) by better understanding the cognitive interview task. Finally, we provide practical guidelines about whom to include as research participants in non-English cognitive interviews. Our recommendations may be also applicable to English cognitive interviews.

Improving Response Rates in Establishment Surveys: Results From Controlled Experiments

Evaluating the Effectiveness of Two Strategies to Improve Telephone Survey Response Rates of Employers

Jeremy Pickreign, *NORC at the University of Chicago*; Heidi Whitmore, *NORC at the University of Chicago*

This paper is an update to work initially presented at the ICES IV conference in Montreal, Quebec, on June 13, 2012. The overall response rate for the California Employer Health Benefits Survey has been hovering between 35 percent and 40 percent since 2004. The response rate varies considerably by certain characteristics, however. For example, the response rate for non-panel firms in 2010 was 24 percent and for firms with 3-49 workers was 27 percent. In contrast, the response rate for panel firms was 61 percent. This study examines two strategies for improving the response rate in surveys of employers by targeting those with the lowest response rate: the smallest non-panel firms. The two strategies included: 1) mailing a personalized advance letter, and 2) offering financial incentives. We pre-called 1,024 non-panel firms with 3-49 workers for the 2011 survey. We sent a personalized advance letter to 513 firms successfully contacted. Simultaneously, we randomly assigned these 1,024 firms into three incentive groups: firms sent a \$20 incentive with the initial mailing; firms promised \$20 upon completion of the survey; and a control group receiving no incentive. Firms sent a personalized advance letter had a significantly higher response rate than those sent a generic advance letter (31.0 percent vs. 18.3 percent, $p < 0.001$). Firms sent a financial incentive with the initial mailing (22.0 percent vs. 28.1 percent, $p = 0.209$) or were promised \$20 upon completion of the survey (30.0 percent vs. 28.1 percent, $p = 0.707$) did not have significantly different response rates compared to firms receiving no incentives. This lack of significance is supported via logistic regression analysis (new work conducted following the ICES IV conference). Sending a personalized advance letter has a significant impact on improving the overall response rate while offering incentives does not.

The Effect of Non-Monetary Incentives in a Longitudinal Physician Survey **Paul Beatty, *National Center for Health Statistics*; Eric Jamoom, *National Center for Health Statistics***

Physicians are often reluctant survey respondents because they are busy and receive many survey requests. The use of incentives appears to be an attractive option for boosting physician response rates, but in some past studies, relatively large incentives (e.g., \$50) have failed to make a difference—possibly because physicians see such incentives as inadequate for “buying” their time. Token incentives may be more effective because they invoke norms of social exchange. In at least one previous study, pens proved to be effective incentives in a general population survey. In this study, we explored whether good-quality pens improved response rates to a mail survey of physicians. We conducted an experiment using pen incentives in the second wave of a three-year longitudinal mail survey, the Physician Workflow Study. This survey, conducted by mail with telephone follow-up for those who do not initially respond, explores physician attitudes and experiences regarding the use of electronic health records (EHRs). The sample was stratified into “adopters” and “non-adopters” of EHRs, then sub-stratified based on their response to the first wave of the survey (early mail respondents, late mail respondents, telephone respondents, and non-respondents). Half of each stratum received a pen incentive with the initial survey mailing. Overall, the response rate for those who received pens was 4% higher than those who did not receive one. Most of the effect was realized in early mail responses. While statistically significant, the main benefit of the boost is that it reduced the need for expensive telephone follow-up. Additional analyses will explore whether the effect of the pen varied based upon responses to the prior survey wave and by EHR adoption experience (as physicians who adopted EHRs might have been more engaged in the survey topic), and whether the pen affected response quality and rates of item-missing data.

Evaluating the Effect of a Non-Monetary Incentive in a Nationally Representative Mixed-Mode Establishment Survey **Manisha Sengupta, *National Center for Health Statistics*; Lauren Harris-Kojetin, *National Center for Health Statistics*; Melissa Hobbs, *RTI International*; Angela Greene, *RTI International***

In 2012, the National Center for Health Statistics (NCHS) launched its new strategy for obtaining nationally representative statistical information about the supply and use of the major types of long-term care providers in the United States—the National Study of Long-Term Care Providers (NSLTCP). NSLTCP represents a substantial redesign, including replacing in-person data collection with less expensive mail, Web and telephone modes. When using in-person data collection over the past couple of decades to survey a variety of long-term care providers (assisted living communities, nursing homes, home health and hospice agencies), NCHS experienced decreasing response rates from highs in the 90s to lows in the 70s. Because of concerns about decreasing response rate trends and achieving adequate response rates when transitioning from in-person data collection to modes that have traditionally produced relatively lower response rates, NCHS embedded experiments into its 2012 national data collection effort. This presentation focuses on a randomized experiment to test the effect of a non-financial incentive. The base protocol included mail and Web choice options with computer-assisted telephone interviewing (CATI) follow-up for non-respondents. The contacts included an advance letter, first questionnaire mailing, thank you/reminder letter, second and third questionnaire mailings, and CATI. For this experiment, treatment cases were offered a tailored report showing their responses compared to all responses, if they participated. We hypothesize that compared

to the control group the treatment group would have a higher response rate both prior to CATI and at study end and lower nonresponse bias, for both provider types. Results and implications for the protocol for the next wave will be discussed.

Examining the Effects of Interventions to Obtain Participation via Less Expensive Modes: Results from Experiments in a Nationally Representative Mixed-Mode Establishment Survey

Lauren Harris-Kojetin, *National Center for Health Statistics*; Manisha Sengupta, *National Center for Health Statistics*; Melissa Hobbs, *RTI International*; Angela Greene, *RTI International*

Decreasing response rates and increasing data collection costs are enduring survey challenges. This presentation reports results from two randomized experiments embedded within two nationally representative establishment surveys, one of 5,000 adult day services centers and the other of 11,700 assisting living communities, both part of the 2012 National Study of Long-Term Care Providers (NSLTCP) sponsored by the National Center for Health Statistics. NSLTCP includes substantially redesigned surveys that changed from in-person to less expensive data collection modes. The main rationale for both experiments is to examine whether survey respondents can be encouraged to participate using less expensive modes. The base protocol for each survey included advance letter, first questionnaire mailing, thank you/reminder letter, second and third questionnaire mailings, then computer-assisted telephone interviewing (CATI) for non-respondents; both mail and Web options were provided in all questionnaire mailings. In the “drive to the Web” experiment, treatment cases were provided only the Web option until the third questionnaire mailing, when they were also given the mail option. Among cases in the “explicit forewarning of non-response follow-up by telephone” experimental group, the thank you/reminder letter stated that if they did not respond via Web or mail by a specific date they would be called to complete the questionnaire by telephone. The premise is that the respondents may prefer to complete the survey on their own schedule, which they can do more readily by mail or Web than by telephone. We hypothesize for both experiments that compared to the control group, the treatment group will have a higher response rate prior to CATI and at the end of the field period. For the Web experiment, we expect a higher response rate by Web compared to the control. Nonresponse bias and item-missing rates will be examined. Results of both experiments and implications will be discussed.

Cell Phone Samples: Effort, Outcomes and Costs

Home Is Where the Cooperation Is: The Association Between Interview Location and Cooperation Among Cell-Phone Users

Christopher D. Ward, *NORC at the University of Chicago*; Becky Reimer, *NORC at the University of Chicago*; Meena Khare, *National Center for Health Statistics*; Carla Black, *National Center for Immunization and Respiratory Diseases*

Interviewing respondents on cell-phones can be particularly challenging to survey researchers. While the proportions of cell-only and cell-mainly households are rising in the United States, cell-phone samples often have lower response rates than landline samples and must therefore sacrifice cost, timeliness, or both. In this paper, we examine the relationship between interview location status (whether respondent is using a landline, a cell-at-home, or a cell-away outside home) and the likelihood to respond. We do so by addressing two questions: first, does

cooperation vary by the interview location? Second, do respondent characteristics interact with the respondent's location on cooperation? We examine these questions using data from the National Immunization Survey, a national, dual-frame RDD survey sponsored by the Centers for Disease Control and Prevention. We use a logistic regression model to investigate factors that predict differences in cooperation and likelihood of break-off among cell-at-home, cell-away, and landline respondents. Preliminary results suggest that time of interview is a strong predictor of likelihood to complete the interview and likelihood of agreeing to release child's healthcare records; evening cell-at-home respondents are much more likely to cooperate than daytime cell-away respondents. Likewise, mothers are more likely to complete the interview and agree to the release of the child's healthcare records than are fathers. Many of these predictors interact significantly with cell-location status. This research provides insight into the behavior of cell-phone respondents and the conditions under which they may be most likely to respond. Given the differences in cooperation among cell-at-home, cell-away, and landline respondents, we will also discuss implications for data quality and limitations of the analysis.

The Cell Effect in Inbound Calling Behavior and Methods for Maximizing Outcomes

Jenny Kelly, *NORC at the University of Chicago*; Becky Reimer, *NORC at the University of Chicago*; Trevor Tompson, *NORC at the University of Chicago*; Jennifer Benz, *NORC at the University of Chicago*

Recent NORC surveys have shown a marked increase in inbound calls that correlates with increasing the proportion of the cell phone sample in RDD studies. The behavior of cell phone respondents is likely to differ from that of landline respondents for several reasons. First, cell phones are unlikely to be listed numbers. Cell users are therefore more likely to expect phone calls from personal or business contacts and a missed call may be viewed as a missed social or business opportunity. Second, cell phones have advanced functionality over many landline phones making it easier to respond to calls—often in a single keystroke. While these factors increase the likelihood of cell phone users to place an inbound call, their expectations of what will occur during the resulting connection may vary from landline users. Understanding how we can best operate within their expectations is critical to obtaining high response rates with cell sample. Using data collected from AP-NORC Center RDD surveys on inbound calls, we analyzed inbound calling patterns to determine the impact of cell sample and methods to maximize good outcomes from inbound calls. Our results confirmed that cell phone respondents were significantly more likely than landline respondents to place inbound calls. Furthermore, we were able to increase positive outcomes when calls were answered immediately by a live interviewer. We also found that inbound cell callers seem mainly interested in simply finding out who called them. As such, we reduced the proportion of hang-ups as respondents waited to be connected by eliminating the automated greeting that allowed respondents to find out the call was about a survey and hang-up before reaching an interviewer. The results suggest that a blended inbound/outbound system can maximize positive outcomes for inbound calls while achieving staffing efficiencies.

Cell Phone Costs Revisited: Understanding Cost and Productivity Ratios in Dual-Frame Telephone Surveys

Thomas M. Guterbock, *Center for Survey Research, University of Virginia*; Andy Peytchev, *RTI International*; Deborah L. Rexrode, *Center for Survey Research, University of Virginia*

An earlier review of a convenience sample found that the per-interview cost of random-digit-dialed (RDD) cell-phone numbers is on average higher than the cost of landline RDD interviews. However, the ratio of cell-phone to landline interviewing costs varies widely across studies and organizations, and may have changed over time. There is reason to believe that the cost ratios for dual-frame phone surveys reported in the 2009 AAPOR Cell-phone Task Force report have become more favorable since those data were assembled. This is especially the case since sampling companies have been improving their frames, data-collection methods have increasingly been tailored to cell-phone samples, and sampling companies have started to provide cell-phone samples with appended information on activity status and ZIP code location of some numbers. The cost ratios also likely vary substantially across data collection designs, such as geographic targeting and screening criteria, and may even vary across sample vendors. We are currently gathering detailed cost and productivity information on a widely inclusive sample of dual-frame telephone studies conducted recently by survey research organizations. We expect relative hourly productivity to depend on whether or not dual-phone users are screened out, the type of cell-phone sample used, the specificity of sample geography, variations in working number density on the cell-phone and landline sides, amount and form of any incentives, variations in interview length, the specificity of screening criteria used in the study, differences in incidence rates, and the efficiency of different dialing technologies. Building on the prior work of the 2009 Task Force, our analysis updates estimates and develops modeling strategies that will allow practitioners to predict more closely the cost of cell-phone calling in future studies by: (1) updating the cost ratio information, (2) expanding the number of surveys and organizations providing input, and (3) identifying the cost ratio drivers.

The Unusual Suspect: Call Protocol and Bias in the 2012 NHTSA Distracted Driving Cell Phone Sample

Paul Schroeder, *Abt SRBI*; Mikleyn Meyers, *Abt SRBI*; Brian Meekins, *U.S. Bureau of Labor Statistics*; Kristie Johnson, *National Highway Transportation and Safety Administration*

To date, the majority of research examining samples of numbers from cell phone and mixed-use exchanges has been limited to response rates, cost estimates, and the reduction in overall bias when combined with landline samples. In the current study we thoroughly examine the call history information in the cell phone sample of a national distracted driving study conducted in 2012. The survey employed a partial overlapping dual frame sample design of households with landline telephones as well as households that relied on cell phones, and collected data from interviews with drivers age 16 and older. The cell phone sample contained 2,143 completed interviews with respondents who were classified as residing in cell-only or cell-mostly households. We review residential penetration; the number of call attempts; the incidence of callbacks, refusals and break-offs; the length of the interview; the time of day of the attempt; and the overall pattern of calling. This allows us to more directly address effort, nonresponse, and bias within the cell phone sample, examining whether increased effort on cell numbers is worth the reduction in bias that may be obtained. We also employ a new technique which allows us to

control for multiple factors and isolate the effect each call attempt has on bias. In addition, our analysis provides recommendations about the calling protocol for cell phones that may lead to increased efficiency when dialing cell phone samples.

A Comparison of Bloomberg Consumer Comfort Index Data in Landline-Only vs. Mixed-Frame Telephone Samples

Julie Phelan, *Langer Research Associates*; Gary Langer, *Langer Research Associates*

The changed nature of telephone use in the United States has raised a quandary for survey research projects that focus on long-term trends in public attitudes. Orthodoxy holds that methodology should be kept constant to preserve these time-trends. Yet changes in access to the sampled population argue for methodological adjustments to preserve coverage. These considerations are acute for ongoing measurements of consumer sentiment, the three longest-standing of which are the monthly University of Michigan/Reuters Index of Consumer Sentiment, The Conference Board's monthly Consumer Confidence Index and the weekly Bloomberg Consumer Comfort Index. The Michigan survey appears to continue as a landline-only sample. The Conference Board in 2011 switched from a mail-in to an Internet-based approach, with admitted impacts on the utility of its time-trend data. Our paper reports on a switch of the third survey, the Bloomberg CCI, from a landline-only to a landline-plus-cell-phone sample in 2012. The change to the Bloomberg CCI was made in recognition of the fact that the number of Americans living in cell-phone-only (CPO) households has reached more than a third of the population. Most survey research firms, finding this level of noncoverage intolerable, have switched to dual-frame designs combining landline-only and cell-phone samples. Given the fundamental value of the Bloomberg CCI's long-term trend, we first conducted an extensive test of the change. From January to March 2012 we supplemented the usual weekly landline CCI (total $n = 2750$) with a weekly cell-phone sample ($n = 1882$). We examined the landline and dual-frame estimates for the weekly index overall, among demographic groups and on individual questions; and assessed the quality of the two samples by comparing their unweighted demographic compositions and design effects from weighting. Apparent differences were tested using overlapping sample t-tests.

Public Opinion on Current Political and Social Issues

Public Support of the Military: Influence of Personal Experience and Perceived Media Coverage on Attitudes Toward the U.S. Army, 2010-2012

Julie L. Andsager, *The Everett Group*; H. A. White, *The Everett Group*; Robert P. Daves, *The Everett Group*; Stephen E. Everett, *The Everett Group*

Public trust and support of the military are crucial as engines of both funding and policy for U.S. military endeavors. This study examines the public's view of the U.S. Army, "through thick and thin," over 11 waves of survey data tracking attitudes quarterly from 2010 through the third quarter of 2012. Data reported comprise responses from 1,950 randomly sampled U.S. adults participating in RDD-based telephone surveys. Nineteen evaluative items regarding Army performance and traits produced two factors (principal-components, varimax rotation), Army professionalism ($\alpha = .92$) and Army responsibility ($\alpha = .88$). Army professionalism comprised items related to training and technological capabilities. Army responsibility included items regarding the care of soldiers, "Wounded Warriors," veterans and Army families, etc. Hierarchical regression analyses indicated that education and age were negatively related to the

attitude that the Army acts responsibly toward its soldiers, their families, and its responsibilities to the nation. Personal experience (including family members currently serving, personal service, and parents' military service) were unrelated to perceptions of Army responsibility. Favorable news coverage was significantly, positively predictive of responsibility, but advertising was not. (Adjusted R2 for full model = .25.) For Army professionalism, age, parents' service, and confidence in the Army were positively related to perceptions of professionalism. Favorable evaluations of news coverage and advertising were also positively related to perceptions of professionalism. (Adjusted R2 = .35.) An examination of evaluations over time for the two indexes indicates some fluctuation, explored in more detail in this paper. This study suggests that, despite polls indicating a decline in news credibility over the last 40 years, news coverage of the Army positively predicts performance evaluations in two aspects. Analysis of comments on news coverage from respondents is included, indicating potential directions for Pentagon Public Affairs professionals attempting to frame news coverage of the Army.

PAPOR Student Paper Award Winner

Too Many Immigrants? Examining Alternative Forms of Immigrant Population Innumeracy

Daniel Herda, *University of California - Davis*

The tendency to over-estimate immigrant population sizes has garnered considerable scholarly attention for its potential link to anti-immigrant policy support. However, this existing innumeracy research has neglected other forms of ignorance, namely under-estimation and non-response. Using the 2002 European Social Survey, the current study examines the full scope of innumeracy for the first time. Results indicate that under-estimation and non-response occur commonly across 21 countries and that over-estimation is far from ubiquitous. Non-responders in particular are found to represent a distinct innumeracy form associated with low cognitive availability and high negative affect. Multilevel models indicate that under-estimation associates with greater opposition to anti-immigrant policy, while over-estimation and non-response associate with greater support. Much of these associations are explained by affective factors. However, significant under- and over-estimation coefficients remain net of controls suggesting that innumeracy may be more important than initially thought. Overall, the results highlight the multifaceted character of innumeracy.

Missed Opportunities in HIV Testing: Health Care Providers Ignore Recommendations and Ignore Seniors

**Micheline Blum, *Baruch College School of Public Affairs, City University of New York*;
Douglas Muzzio, *Baruch College School of Public Affairs, City University of New York***

'Missed Opportunities' reports on a survey conducted by Baruch College Survey Research for the NYC Department of Health and Mental Hygiene. The study surveyed 2473 adult New Yorkers from June to August, 2011 on a variety of HIV-related and sexual behavior questions. A 2012 AAPOR paper reported some preliminary findings from this study. "Missed Opportunities" reports on analysis, data and policy implications NOT presented in 2012. Background: In 2006, the CDC recommended routine HIV screening for individuals aged 13-64 in healthcare settings. A 2010 New York State (NYS) law mandated the offer of an HIV test to all patients aged 13-64, receiving hospital and primary care services. New Yorkers aged 65+ accounted for 2% of new HIV diagnoses in 2010. However, 47% of them were diagnosed late in the course of infection, more than double the rate of the general NYC population (22%). Findings: 1. Health care providers are ignoring both the NYS law and the CDC recommendation to offer an HIV test to all

patients aged 13-64 (93% not offered test) with profound health consequences; 2. Seniors (65-74) are particularly apt to have never been tested for HIV and to not be offered a test by health care providers, despite the fact that one in three aged 65-74 is sexually active. Consequences and Policy Recommendations: If NYC health care providers adhered to CDC recommendations and NYS law, nearly a million New Yorkers 18-64 would have been tested for HIV for the first time, leading to discovery of about 6500 with previously undiagnosed HIV infection. Among those aged 65-74, an estimated 200,000+ would be tested for the first time, again uncovering previously undiagnosed HIV infection and decreasing the number of late HIV diagnoses.

What Explains California's Passage of Proposition 30: Fear of Education Cuts, Gubernatorial Approval, Political Trust, or Tax Preferences?

Dean E. Bonner, *PPIC*

With automatic trigger cuts—mostly to K–12 education—looming, Californians went to the polls on November 6 and passed Proposition 30, which increases taxes on the wealthy for seven years and the state sales tax by 1/4 cent for four years. Given the central role that Governor Brown played in the initiative campaign this paper analyzes the role that gubernatorial approval—in comparison to opposition to automatic education cuts, political trust, and tax preferences—played in support for Proposition 30. Utilizing pre-election data of those most likely to vote in the general election, preliminary results indicate that tax preferences exert the most leverage on Proposition 30 support. Further analysis will examine the interplay between support for Proposition 30 and support for Proposition 38, another measure to fund education that relied on increasing income taxes on most Californians based on a sliding scale. Analyzing what types of people supported both propositions will provide a broader understanding of support for tax increases in California.

Racial Resentment, Belief in Rumors about Barack Obama, and Racial and Ethnic Identities

Michael W. Traugott, *University of Michigan*; Ashley E. Jardina, *University of Michigan*

Barack Obama has been the focus of innumerable rumors about his citizenship and religion, and recent research has shown that racial resentment plays an important role in explaining these views. These analyses have been based almost entirely upon white survey respondents because the measurements of these concepts were made in single cross-sectional surveys, and the numbers of nonwhite respondents were too small for analysis. In this analysis, we employ a unique multi-survey data collection that allows pooling of respondents to support separate analyses of Black and Hispanic respondents in addition to whites. As a result, we compare models explaining these beliefs among different segments of the population and discuss why and how they differ. This analysis is complicated by lower acceptance of rumors about Obama among black respondents.

Reaching and Estimating Small or Specialized Populations

Dynamic Averaging: A Modified Time Series Approach to Improve Estimates for Smaller Demographic Groups

Kelly Dixon, *Arbitron*; Al Tupek, *Arbitron*; Richard Griffiths, *Arbitron*; Wolfgang Jank, College of Business, *University of South Florida*

Arbitron is a media research company that produces quarterly or semi-annual estimates of radio listening. The surveys are designed to produce a certain standard error for estimates for the radio-market at an aggregate level. However, our primary customers are radio stations that target specific demographic sub-groups of the market (for example, Black Males 18-34.) The standard errors on these smaller sub-group estimates of radio listening are large which makes it difficult to see a trend in the estimates. Our research goal is to improve estimates reliability for smaller sub-groups and geographies while retaining trends and other actual changes of the more reliable aggregate estimates. We also need a solution that will allow for sub-groups to add up to the aggregates. Our proposed solution, dynamic averaging, achieves a smoother time series for the sub-group estimates, which should give customers a better view of long-term trends in their estimates. We estimate the reliability improvements to be equivalent of a sample that is two to three times as large.

Small Area Estimation of a Rare Population Incidence

Stanislav Kolenikov, *Abt SRBI*; Benjamin Phillips, *Abt SRBI*

Abt SRBI is conducting a large scale CATI Survey of American Jews. Identifying this rare population with incidence of about 1.9% across the nation involves enormous screening effort. As Kalton & Anderson (1986) suggested, stratifying this population into different levels of incidence brings efficiency gains in both costs and accuracy. Thus to inform stratification and coverage decisions for the sample design, reliable estimates of Jewish incidence were needed. We developed a small area mixed effects model that combined information from a number of sources. The unit-level dependent variable, 0/1 indicator of being Jewish by religion (JBR), as well as the sampling weights, came from a merge file of national studies conducted by Pew Research Center. The area (county) level data came from: 1) ICPSR County Characteristics data set; 2) a commercially acquired list of synagogues (geocoded by Abt SRBI GIS unit); 3) a list of Jewish educational organizations (JData.com at Brandeis University); and 4) incidence of Jewish names by a commercial sample provider. We demonstrate the steps of fitting the multilevel mixed effects logistic regression model, obtaining the empirical best predictions (EBP; Jiang and Lahiri 2001), and using the EBPs to delineate the strata by incidence and estimate undercoverage of the survey. We also provide qualitative comparisons of our SAE estimates with alternative estimates, such as direct estimates based on the screeners for the survey, the direct estimates based on merge file only, and incidence of Jewish names only.

Efficient Sampling Designs for Rare Populations

Benjamin Phillips, *Abt SRBI*; Stanislav Kolenikov, *Abt SRBI*

Abt SRBI is conducting a national dual frame RDD Survey of American Jews. As a rare population (c. 1.9% of adults), a very large proportion of the study budget is spent identifying eligible households, necessitating the development of an efficient sample. The methods we

develop apply to other rare populations. Study objectives included minimizing sampling error for estimates of total Jewish population incidence and for estimates of characteristics of the overall Jewish, Orthodox, and Russian Jewish populations. Given different geographic distributions of these groups, targeting these populations had different implications for sample design. Design choices included sample size, allocation to strata, landline/cell allocation, and degree of undercoverage. Using small area estimates of Jewish incidence at the county level (Kolenikov and Phillips, submitted) we developed an optimal allocation using the Excel 2010 nonlinear solver and the study design/budget spreadsheet. Study objectives—measured as effective sample sizes for the above subpopulations of interest—were weighted by importance and combined using the Cobb-Douglas function. Design effects were calculated by simulating survey dispositions based on a “donor” survey expected to have similar outcome rates. This allowed us to include the impact of landline/cell phone RDD allocation on design effects, including frame integration and adjustment to NHIS coverage estimates. We were also able to estimate design effects for subpopulations of interest. It was determined that a smaller than initially planned sample yielded a greater effective sample size, with the reduction in sample size allowing the use of a more efficient design under the study budget. We illustrate the effect on sample design of varying the importance of study objectives, coverage of the Jewish population, and enforcing a minimum sample size constraint. We compare design projections to study results and, in addition, compare study design to field results, such as design effect and screening rate.

Sampling “Hidden” Populations in Developing Countries: An Application of Respondent-Driven Sampling (RDS) in Ethiopia

Charles Q. Lau, *RTI International*; Georgiy Bobashev, *RTI International*; Burton Levine, *RTI International*

In developing countries, surveys often attempt to sample populations not readily accessible to researchers such as drug users, sex workers, or entrepreneurs. For example, smaller businesses oftentimes lack a fixed address or operate out of plain sight, meaning that sampling frames for these businesses are typically unavailable or incomplete. As a result, typical methods such as sampling from telephone directories suffer from substantial coverage errors. One potential solution is respondent-driven sampling (RDS), a method similar to chain-referral or “snowball” sampling but with strict controls over the number of people a single subject can recruit. Mathematical theory behind RDS allows one to adjust for known biases in convenience sampling, such as network homophily and the tendency for individuals from larger networks to be overrepresented. RDS has been used productively in studies of many hidden populations, but has not been applied to businesses. To address this gap in our understanding, we used RDS in a survey of businesses in the capital city of Ethiopia, Addis Ababa. Eligible participants were owners of small businesses (3-99 employees) in the manufacturing, service, or trade sectors. We recruited 24 initial respondents non-randomly. These initial respondents (and all subsequent respondents) were provided incentives to recruit up to 3 additional respondents. After 11 waves of recruitment over six weeks, we achieved our target sample size of 608. Our paper reports the statistical properties of our sample and critically evaluates the assumptions underlying the RDS approach. Preliminary analysis shows the sample compositions converged and reached equilibrium which, according to the theory, indicates representativeness of the population. The sample composition also closely tracked to the estimated population composition, suggesting that RDS could be used as a reliable but also a cheaper alternative to probability sampling. We also discuss the levels of homophily and random recruitment of the network members.

Issue Publics in Nanotechnology in the New Media Environment

Doo-Hun Choi, *University of Wisconsin – Madison*; Michael Cacciatore, *University of Wisconsin – Madison*; Young Mie Kim, *University of Wisconsin – Madison*; Dietram Scheufele, *University of Wisconsin – Madison*; Michael Xenos, *University of Wisconsin – Madison*; Dominique Brossard, *University of Wisconsin – Madison*; Elizabeth Corley, *Arizona State University*

While the public has low levels of interest in and understanding of nanotechnology (Scheufele & Lewenstein, 2005), instead, issue publics, subsets of the population who are passionately concerned with a specific science issue (Kim, 2009), exert significant influences on public opinion formation and science policy decision. The concrete attitude structure and issue specializations of issue publics indicate a high level of attitude extremity. Given the preferences of selective information attendance and the availability of diverse and specialized information, the nanotechnology issue publics will likely use the Internet to strength their attitudes. Analyzing nationally representative online survey data (N = 2,805) collected by KnowledgeNetworks, this study explores the predictors of attitude formation among the nanotechnology issue publics. This study also explores how these factors interact in determining attitude extremity toward the nanotechnology in the new media environment. Our findings showed that “nanotechnology issue publics” tend to use science media more attentively and have a higher level of nanotechnology knowledge than non-issue publics. The issue publics were more extreme in their attitudes toward this emerging technology. We also found that the Internet contributes to an increase in attitude extremity among issue publics. More importantly, exploring how the issue publics form their attitude extremity, the study found that issue publics relied on their schema, a specific cognitive structure toward a certain issue, for science and technology, rather than political ideology, a set of basic beliefs about political world. Given opportunities for selective exposure online and the interpretative schemas, the issue publics can be expected to become polarized and extreme when forming their attitudes toward nanotechnology. In this sense, as informal opinion leaders, the issue publics will have the potential to influence public opinion formation toward nanotechnology, mirroring the extreme division between the issue publics.

Monitoring Interviewer Behavior

Detecting Poorly Conducted Interviews

Joerg Blasius, *University of Bonn*

In their recent book, Blasius and Thiessen (2012) introduced several screening methods to assess the quality and validity of survey data. They characterized the survey interview context as one in which task simplification, time and effort minimization, and cost reduction strategies by respondents, interviewers, and research institutes resulted in poor data quality. In this presentation, we concentrate on the quality of the interviewers, identifying patterns that help to assess how carefully and thoroughly they conduct their interviews. We illustrate our ideas using the German General Social Survey 2008 in which we detect clusters of interviewer-specific response combinations whose frequency of occurrence defies the odds to such an extent that we suspect interviewer fraud to be the cause of some of them. Using two of the screening methods proposed by Blasius and Thiessen (2012) we find a substantial number of interviewers who simplified their tasks in a manner that reduced their interviewing time and effort but

increased their “measurement error”. Blasius, Jörg and Victor Thiessen (2012): *Assessing the Quality of Survey Data*. London: Sage.

Interviewer Affect and CARI Effects: Lessons in Implementation and the Effects of CARI on a Large-scale Longitudinal Study

Ryan A. Hubbard, *Westat*

This paper summarizes the evolution of Computer Audio-Recorded Interviewing (CARI) on Medicare Current Beneficiary Survey (MCBS). CARI is integrated into a number of studies due to its value as a tool for interview validation, assessment of interviewer performance, evaluation of data quality and question assessment. While these uses of CARI have been documented (Biemer, 2000, 2003; Herget, 2001; Thiessen, 2007, 2009; Fisher, 2012; Kinsey, 2012; Hicks, 2008, 2012), CARI implementation on MCBS offers a new perspective on the effect CARI can have on operations when implemented on an ongoing longitudinal study with a rotating-panel design. MCBS implemented CARI with experienced interviewers and longitudinal study respondents accustomed to an interview for which audio recordings never occurred. Interviewers voiced concern that introducing CARI would negatively affect rapport with respondents. In fact, the project initially experienced a high rate of refusal to record concentrated among a subset of experienced interviewers. Efforts were made to improve consent during the field period, and the refusal to record rate dropped to expected levels. MCBS later introduced new interviewers and respondents who had not been conditioned regarding CARI. The comparison of these combinations contributes to a better understanding of CARI effects. Both interviewers and respondents conditioned to a non-recorded interview are expected to produce higher rates of recording refusal, but the key to this effect lies in the interaction. While implementing CARI on an established study offered insight into consent, MCBS study design also allows for a data quality impact assessment. A 10 minute increase in average interview length on MCBS was attributable to CARI and the enforcement of proper interviewing technique (fewer shortcuts, reading full question text). This practice should lead to better data quality. The paper analyzes the effect of CARI on the quality of event reporting through a comparison of respondent-reported medical events to insurance statements.

Variability in Error Detection Among Telephone Monitors

Douglas B. Currivan, *RTI International*; Derek Stone, *RTI International*; Curry Spain, *RTI International*; Nicole Tate, *RTI International*

Standardized methods and tools for monitoring telephone interviewers are important for ensuring survey data meet high quality standards. In order to effectively limit the risk of interviewer behaviors biasing or adding variance to survey estimates, the quality monitoring process requires accurate and consistent detection of interviewer errors. To this end, RTI has developed a standardized, mode-independent interview quality monitoring evaluation system, QUEST. This system supports evaluation of interviewing quality through both live monitoring and review of digitally-recorded sessions. QUEST allows telephone interviewing behaviors to be evaluated using a common set of quality metrics that are stored in a single shared database. These metrics are based on objective indicators of specific interviewer behaviors, including definitions and concrete examples for each behavior, as opposed to more subjective ratings or impressions of interviewing quality. The primary hypothesis of our research is that the standardized, objective approach followed in QUEST will produce minimal variation across monitors in their detection of interviewer errors and other unacceptable behaviors. Two primary

sources of data are used to investigate variability in the rates at which monitors detect interviewer errors: comparison of error detection rates across monitors from monthly monitoring results and examination of the results of blind scoring by monitors of a set of 10 selected interviewing scenarios. Comparisons of error detection rates across monitors includes both overall errors detected across sessions and errors detected for specific interviewing skill areas. In addition, this analysis examines whether scoring across monitors varies when factors such as interviewing shifts or monitor experience levels are considered. Based on the results of the comparisons of monthly monitor scores and blind scoring of interviewing scenarios, this presentation discusses the implications of the observed levels of monitor scoring variability in general and disagreements on specific scenarios for accurate and consistent detection of interviewing errors.

A Field Experiment Using GPS Devices to Monitor Interviewer Travel Behavior **Kristen Olson, *University of Nebraska-Lincoln*; James Wagner, *University of Michigan***

Survey organizations rely on interviewers to make informed and efficient decisions about their efforts in the field, including which houses they approach to knock on doors, make appointments, and obtain interviews (Groves and Couper 1998). Previous evidence suggests that inefficient decisions about where to travel can have deleterious effects on response rates (Wagner and Olson 2011). To date, however, there is no systematic evaluation of how interviewers make travel decisions in real time. This paper presents initial findings from a field experiment and a survey of interviewers in a face to face survey, the National Survey of Family Growth. NSFG interviewers were equipped with GPS-enabled smartphones. In the first quarter, a random half of the interviewers were asked to record their travel behavior via a GPS logging app in the smartphone; the second group recorded their travel behavior during the second quarter of data collection. All interviewers were asked to record their travel for subsequent quarters. We evaluate interviewer compliance with the GPS request, the quality of the recorded GPS points, the correspondence between the GPS points and the attempts recorded in the call records, and provide an overview of the interviewers' travel behavior. We also report results from a survey of the NSFG interviewers about the smartphone and GPS logging app. Initial results indicate that 68% of the first quarter interviewer-days, 54% of second quarter interviewer-days, and 57% of third and subsequent quarter days had GPS data recorded. Results from the interviewer survey indicate that an interviewer's failure to have travel behavior recorded resulted largely from technical problems (e.g., forgetting to turn the phone on), not from discomfort with having movements tracked via the GPS device. Implications for future use of GPS devices to monitor interviewer travel behavior will be discussed.

Friday, May 17
3:15 p.m. – 4:15 p.m.
Poster Session 2

1. Trends in Cell Phone Calling Outcomes: BRFSS 2008-2011

Carol Pierannunzi, Centers for Disease Control and Prevention; Machell Town, Centers for Disease Control and Prevention; Simone Salandy, Northrup Grumman Contractor for CDC; Lina Balluz, Centers for Disease Control and Prevention

In 2011, the Behavioral Risk Factor Surveillance System (BRFSS) released both landline and cell phone data for public use for the first time. However, the BRFSS has collected cell data since 2008 as part of a large pilot study. This study examines the calling outcomes for 2.7 million cell phone numbers included in the BRFSS samples from 2008-2011. Trends in final dispositions are examined over time for the aggregated state samples and for selected individual states with large number of cell samples. Patterns of response rates, refusal rates, contact rates, out of sample numbers, terminations and partial completes are illustrated. Demographic characteristics of respondents who completed the screening questions are also included. Four year trend lines are produced for interim calling outcomes resulting in completed interviews as well as for calls which result in refusals or cut-offs. Results indicate that although terminations, break-offs, partial completes and refusal after determination of eligibility are relatively small percentages of the sample, the proportion of these outcomes is increasing over time. When taken as a percentage of the sample which resulted in contact with potential respondents, these trends in unsuccessful cell phone outcomes are more pronounced. The BRFSS is currently conducting new pilot studies to determine the feasibility of other modes of data collection to counteract these trends.

2. Non-Responds Reasons Among Surveys Participants in the Gulf Arab Countries, Case of QATAR

Elmogiera Elawad, Social and Economic Survey Research Institute, Qatar University; Mohamed Ahmed Bala Agied, Social and Economic Survey Research Institute, Qatar University

Choosing appropriate times of interviewing and use of males and females interviewer's to meet the society customs in addition to translate the questionnaire's into different languages, all these procedures taken into consideration to improve or maintain the response rate in Qatar households surveys, but this rate obviously declined. With each field survey conducted by the Social and Economic Survey Research Institute at Qatar university -SESRI, we asked selected participants who refused to participate in the study about the reasons of rejections if any, some of them of course refused even to answer this question, but some answered. In this paper we will try to understand non respond cause among Qataris and expatriates live in Qatar state, by studying the answers of non-respondent's in SESRI previous surveys 2011- 2012, we will know the reasons of non-responds and the attitude of Qataris and expatriates towards the participations in field surveys.

3. Internet Versus Mail: A Comparison of Data Quality Indicators

Jenifer G. Tancreto, U.S. Census Bureau; Rachel Horwitz, U.S. Census Bureau; Mary Davis, U.S. Census Bureau; Mary Frances Zelenak, U.S. Census Bureau

In April 2011, we conducted a test to evaluate the feasibility of providing an Internet response mode to households selected for the American Community Survey (ACS). The main purpose of this test was to determine the best methods for informing people about the Internet response option and encouraging them to respond. Results suggested that

providing a sequential mode offering, starting with Internet followed by a paper questionnaire, maintained or increased response rates while driving over 50 percent of self-response to the Internet (Tancreto et al., 2012). This study analyzes data collected in that test, as well as supplemental data collected as part of a reinterview, to examine the quality of the data collected on the Internet compared to the quality of mail response data. This analysis will help determine whether the Internet provides data of comparable quality to mail. Specifically, we used the following data quality indicators: the amount of outliers and the percentage of rounded values for some numeric income fields; the correlation between certain related measures; and measures of response error generated from comparing data from the original interview to a reinterview. We attempted to control for known demographic differences between mail and Internet respondents using propensity weights so we could measure true mode differences. Overall results suggest that the Internet data appear to have a comparable level of quality compared to mail data.

4. Reducing Erroneous Enumerations in the Decennial Census Group Quarters Populations While Potentially Reducing Follow-Up Costs

Geoffrey Jackson, U.S. Census Bureau

The foundation of the decennial census is to successfully count each person once, only once, and in the correct place. Sometimes people live or stay in more than one place and their lengths of residency make it difficult to ascertain which place is correct. The Census Bureau has a rule on where people should be counted; how the rule is applied depends on if the person lives in a housing unit or group quarters (living quarters such as college dormitories, prisons, etc.) and where they were living on April 1, 2010. The respondent is not always aware of how the rule applies to their situation. Research has shown that some people living in group quarters tend to also be counted at a housing unit. In the 2010 Census, the method for resolving this person duplication occurred if 1) on the housing unit questionnaire a person indicated living or staying at another address, 2) that housing unit was re-contacted and the duplicated person was removed during the costly Coverage Followup Operation. During the 2010 Census, an experimental questionnaire was tested for people living in group quarters. The experimental group quarters questionnaire asked all respondents if they had another address where they stayed at besides the group quarters. The traditional group quarters questionnaire only asks for another address if the respondent indicated they lived or stayed somewhere else most of the time. This paper will analyze the number of people that provided an address on both group quarters questionnaires and if they were found to be counted at that those addresses. The paper will show that the person duplication between a group quarters and housing unit can be resolved without any costly follow-up interviews by using the collected address on the questionnaire in conjunction with the results of the person duplication matching.

5. Attempting to Reduce Respondent Burden in Complex Listing Tasks

Lauren A. Walton, The Nielsen Company; Anh Thu Burks, The Nielsen Company; Christine Pierce, The Nielsen Company

In order to gain survey participation, researchers try to make the benefits of participating outweigh the drawbacks of not participating by offering incentives, shorter questionnaires, or interesting survey topics. Each survey is unique in its level of potential respondent burden moderated by the questions asked, the survey format, and the level of cognitive effort required by a respondent to complete the survey. When designing a survey, one major consideration to gaining cooperation is the amount of time and effort required of a participant to complete the survey. This paper tackles the respondent burden associated with a complex knowledge based listing task that can be arduous to complete. An experiment was conducted using a paper and pencil survey where respondents were asked

a series of demographic questions followed by a complex knowledge based listing task (i.e., respondents can provide hundreds of specific pieces of information) finished with the respondent completing an event history calendar of a specific activity. More specifically, the experiment manipulated the listing task that a respondent was asked to complete in the survey. A proportion of respondents were randomly selected to provide a detailed list of information while others in the sample were assigned to provide the minimal amount of data required in order to reduce respondent burden associated with the listing task. Preliminary results indicate a highly significant difference in favor of the reduced listing task in the number of households that returned a useable survey (18.5% vs. 17.9%). Results from this July 2012 test suggest reducing respondent burden in challenging surveys is a good thing for respondents and research organizations.

6. Predicting Biases Due to the Use of Lottery Incentives in Surveys

David Fan, *University of Minnesota*; Joe Murphy, *RTI International*; Susan Mitchell, *RTI International*; Ken Blake, *Middle Tennessee State University*

The goal of a survey is to obtain a set of responses from a representative sample of a target population. Typically defined, representativeness means the characteristics of the sample will, on average, match the target population. In other words, the survey methodology must be independent of the responses sought. For example, the telephone method is commonly used for political polls under the assumption that the responses are independent of phone usage. However, the same phone poll would not be used to determine why the respondent does not use the telephone. The reason is the complete correlation and lack of independence between the phone non-usage question and the survey mode. The phone poll example shows how error may lead to representative responses to some questions but not others. This paper explores a similar inquiry, but about bias due to choice of incentive type. Specifically, do lottery or drawing-type incentives lead to biased data for certain types of questions? To identify the potential effect, we included a question about the preference for a lottery incentive on two separate surveys using either a fixed payment incentive or no incentive. We asked whether the respondents would prefer a drawing to a fixed payment. We scored for the independence of lottery response from responses to other survey questions. Responses correlated with a lottery preference should not be used in surveys with a lottery incentive because the lottery is likely to bias the results. This paper is a demonstration project for identifying potentially problematic questions on surveys. The long term goal is to encourage survey researchers to routinely add simple methodological questions like this to surveys. A database could be constructed for the types of responses that are correlated with various survey designs and hence be problematic using the corresponding methods.

7. Tell Me the Truth: The Response Validity of College Student Populations

Cole Napper, *RTI International*; Tilman Sheets, *Louisiana Tech University*

According to Peterson's (2001) meta-analysis, a considerable proportion of research in the social sciences has been conducted using American college sophomores as participants; also known as the "science of sophomores" (Gordon et al., 1986). Although some researchers support the notion that undergraduate students can be representative populations for generalization to non-student populations (Highhouse & Gillespie, 2009), this assertion should be evaluated in the context of whether participant's motivation is to satisfice or provide accurate responses (Krosnick, 1999). Fan (2006) states about half of what participants report on self-report questionnaires is inaccurate. This is a troubling finding for social scientists, and should prompt researchers to assess the quality of their data before they expand upon their research conclusions. This research study was conducted to assess response validity of an undergraduate student population. An experimental design utilizing

deception was used to elicit truthful responses on the effort and motivation of students completing a long self-report questionnaire. The purpose was to examine if undergraduates' responses in survey research are dishonest, involve little or no effort by the participant, and if participants intentionally provide inaccurate responses. After finishing a cumbersome 300-item scale, participants completed a response validity scale (RVS) which indicated the level of effort they exerted and whether they intentionally provided inaccurate responses to the self-report questionnaire. However, while participants completed the RVS, they were told they were being monitored for lie detection (i.e., inactive eye-tracking and EEG hardware were used to create a ruse that untruthful responses were monitored). The results examined the validity of using long psychological measures (i.e., 300 items) on college student populations. Also, student responses to the RVS are discussed, as well as the relationships between those students who failed the validity check and the students who admitted to intentionally providing inaccurate responses.

8. Utilizing GIS Data to Enhance Survey Data

Christine Cowles, *Abt SRBI*; Mark Morgan, *Abt SRBI*

Researchers have an increasing number of non-survey data resources available and it is essential that the survey research community is proactive in incorporating this added value in their study designs. This methodological brief will examine the use of geocoding and appended geographical statistics in the analysis of how one's neighborhood can impact their mental health. The aim of the research is to understand the role of neighborhood environment on physical and mental health to encourage policy choices that improve the opportunity for aging residents to avoid or minimize depression and its effects on quality of life. The data are collected in a 3 year cohort survey conducted among 3,500 older residents living in New York City.

9. The Impact of Climate Change Issue in the 2012 U.S. Presidential Election

Bo MacInnis, *Stanford University*; Jon A. Krosnick, *Stanford University*; Jon Cohen, *Capital Insight/Washington Post Media*; Clifford Young, *Ipsos*

The long-held theories of voting behavior posit that voters evaluate political candidates on the basis of their positions on issues, and yet have received little empirical confirmation in the general population and limited support among members of the public who attach high personal importance to the issue. National surveys show large majorities of Americans believe in climate change and want government actions to reduce future climate change, and that climate change issue public is sizeable, suggesting climate change would be an important factor in the 2012 election. However, counterarguments exist: one is that other issues such as economy, seemed prominently important to electoral, possibly diminishing the importance of climate change and the other is that candidates appeared nearly silent on climate change, leading to the tendency toward a null effect of issue voting. Based on the data from nationally representative surveys in September 2011 and June 2012, this research employed the well-established methodologies in political science through the measure of issue congruence. In the first study where respondents chose candidates to vote for if the election were to be held where the candidates were President Obama and one named Republican candidate, we exploited the cross-candidate and cross-respondent variations in climate change stances as the identification sources, and found that Americans were more likely to vote for a candidate, Democratic or Republican, whose belief matched their own than to vote for a candidate whose belief differed from their own on climate change. The second study found that greater relative proximity to Mr. Obama on climate change than to Mr. Romney increased the likelihood of voting for him instead of for Mr. Romney. While issue voting was found to be present among the general population in both

studies, it was moderated by attitude strength and personal importance, consistent with issue voting theories.

10. A Framework and Usage Model of Social Media for Young Adults

Jennifer C. Romano Bergstrom, *Fors Marsh Group*; Caitlin Krulikowski, *Fors Marsh Group*; Ricky Carroll, *Appalachian State University*; Kara Marsh, *Fors Marsh Group*; Joseph N. Luchman, *Fors Marsh Group*; Katie Helland, *Joint Advertising, Market Research & Studies (JAMRS)*; Megan Fischer, *Fors Marsh Group*

The use of social media has grown immensely over the past decade, with technological and Internet innovations like Facebook, Twitter, and YouTube achieving massive adoption in a few years. Increasing numbers of young adults are using social media, and many companies and organizations are using social media to reach out to youths. However, it is unclear as to the extent to which organizations can apply the same strategies across products, services, and industries when introducing social media into their marketing strategy. Moreover, it is unclear whether social media is the most viable or effective marketing platform to reach out to all young adults. We reviewed existing social media literature (e.g., popular press, academic journals). Our review reveals that little guidance exists on how and why youths use social media. Our review was used to build an organizing framework of social media usage that was subsequently tested using a national probability-based pencil-and-paper survey (N = 3,743) and a follow-up Web-based survey that the original respondents were invited to complete (N = 1,686). Data for the original survey were analyzed using a finite mixture model approach to uncover underlying “classes” of social media user profiles. Data for the follow-up survey were analyzed using multidimensional scaling to uncover the underlying framework across myriad social media channels. We present the resulting two-dimensional framework model and the usage model, which demonstrates the way young adults use each type of social media (e.g., “pushing” information; “pulling” information). Our results suggest successful social media strategies depend on the function of the social media channel and the marketing objective. Most importantly, our study provides critical information on the motives of the media user in using social media—critical information for effective targeting. We conclude with recommendations for organizations seeking to use social media for marketing efforts.

11. Surveywalls: A Breakthrough for Survey Customers or DIY Run Amok?

Tom Wells, *The Nielsen Company*; Elizabeth Dean, *RTI International*; Kumar Rao, *The Nielsen Company*; Joe Murphy, *RTI International*; David Roe, *RTI International*

Online surveys continue to transform how survey research is conducted, not just in terms of the capabilities they offer, but also how online surveys are designed. Several companies have recently entered the survey research field with a new type of platform, offering researchers a do-it-yourself (DIY), cost-effective approach to surveying thousands of people online. Respondents to DIY surveys are recruited from an online panel of Internet users or by using a variety of online recruitment methods, including banner advertisements, email campaigns, and search campaigns (i.e., search engine generated links). A new recruitment approach for conducting DIY surveys has been gaining traction -- a “surveywall” that first intercepts Internet users attempting to access restricted/paid content from a participating website then solicits them to participate in a very brief survey (1-2 questions). Users are sampled in real-time and, in exchange for their survey participation, are given access to the paid content. Proponents of this DIY approach argue that by reducing survey burden, and simultaneously providing more meaningful incentives (i.e., access to content), survey results are as accurate as those derived from probability-based online panels. In this study, we test the feasibility and performance of intercept-type DIY survey relative to a probability-based online panel, a traditional opt-in online panel, and online populations recruited through two

popular social media platforms using a common questionnaire. We provide an independent assessment, useful to those studying and contemplating using such a system. We compare responses from all platforms to demographic and behavioral benchmarks, using the average percentage point absolute error across all the questions in the survey, as done by McDonald, Mohebbi, and Slatkin (2012) and Yeager, Krosnick, et al. (2011) in their comparative research on survey accuracy. We discuss the findings from the study and conclude with a call and recommendations for further research on this topic.

12. Does Classroom Observer Reliability Differ By Content or Approach To Data Collection?

Harshini K. Shah, *Mathematica Policy Research* ; Jillian Stein, *Mathematica Policy Research*; Katherine M. Burnett, *Mathematica Policy Research*; Tim Bruursema, *Mathematica Policy Research*

The use of classroom observations (CO) has become increasingly common in large-scale education studies assessing teaching effectiveness and in state accountability systems. COs capture the actual experiences of students in classrooms rather than the intended instruction that is often captured through teacher reports. COs require observers to record behaviors in the classroom and can range from structured checklists to more qualitative descriptions of behavior. A major consideration when using COs is observer reliability. Although it is well understood that low observer reliability has an adverse effect on the quality of data, surprisingly little research has compared the reliability of different approaches to reporting live CO data. Our paper draws from a large-scale study that used a CO tool in first and second grade classrooms. The tool used in this study contains a combination of items that involve 1) discrete behavior coding and 2) global ratings of classroom processes. The same observer completed both types of ratings for each CO and observers collected data in classrooms in more than one region. We compare the reliability of these two approaches by decomposing the variance in each type of rating across regions, classrooms, and observers to determine how much of the variance in scores was attributable to the observer and how this varies by observation approach (behavior coding versus global ratings). Because the content of items differs both within and across the two approaches, we also examine the extent to which content influences reliability. Our findings contribute to the field's understanding of issues surrounding the selection and development of observational tools and facilitate the collection of higher quality data when using COs.

13. An Application of Network Analysis for Mapping the Structure and Evolution of an e-Journal

Kumar Rao, *The Nielsen Company*; Kirby Goidel, *Louisiana State University*; Ashley Kirzinger, *University of Illinois Springfield*; John M. Kennedy, *Indiana University*

Over the past decade, the landscape of scholarly publishing is changing the way journal information is disseminated. While most journals nowadays offer digitized version of their articles available for access on the Web, some have taken a step further by publishing online-only articles, rather than in print. One such electronic journal, or e-journal, is Survey Practice (SP). Established in 2008, the mission of SP is to provide current information on issues in survey research and public opinion that is useful to survey and public opinion practitioners, new survey researchers, and those interested in survey and polling methods. The articles in Survey Practice emphasize useful and practical information designed to enhance survey quality by providing a forum to share a) advances in practical survey methods, b) current information on conditions affecting survey research, and c) interesting features about surveys and people who work in survey research. In this study, we use sophisticated network analysis techniques to map the structure and evolution of SP over a four year time frame. In this effort, we formulate and investigate the following few research

questions: How has the collective scholarly knowledge of SP grown over time? Are there any trends that are shaping the overall knowledgebase in SP and is this in line with SP's mission? What major areas and topics of survey and public opinion research have been addressed in SP, and how have they evolved across years and how are they interlinked to one another? Are certain segments of published articles (such as sponsored/funded research) in SP evolved differently than the overall journal when it comes to intersection with certain popular research themes and topics? The answers to these questions would provide a basis to study the impact of SP in promoting research on issues in survey research and public opinion.

14. Who Knows: Question Format, Don't Know Discouragement, and Estimates of Political Knowledge as a Dependent and Independent Variable

Joshua Robison, *Northwestern University, Political Science Department*

Political knowledge is a key dependent and independent variable. However, there is considerable debate concerning how best to measure this concept, particularly as it relates to open-ended versus multiple choice formatted questions and the encouragement of don't know responses. I report data from a survey experiment contained on the ANES' EGSS3 survey, conducted in December 2011, which sheds further light on these questions. The evidence supports three key findings. First, discouraging don't know responses via an introductory text do not substantially increase estimates of political knowledge. Second, opting for a multiple-choice format rather than open-ended questions has a substantial impact on estimates of how knowledgeable the mass public is and who is knowledgeable. Purported knowledge gaps based upon political interest, education, and race are all ameliorated when using a multiple-choice knowledge question rather than an open-ended one. Third, these design elements influence the purported relationship between knowledge and stereotype holding. While knowledge is negatively related with stereotype holding against African-Americans, Hispanics and Muslims when an open-ended format is used, this relationship is null when estimates of knowledge from multiple-choice questions are used instead. The results reported in this article are thus highly relevant for the measurement and use of a critical variable in political analyses (knowledge about politics).

15. The Results of Usability Testing of a New Online Consumer Expenditure Web Diary
Kathleen T. Ashenfelter, *U.S. Census Bureau*; Marylisa Gareau, *U.S. Census Bureau*

Two rounds of usability testing were conducted on a prototype version of the Consumer Expenditure (CE) Web Diary Survey from January-July 2012. The CE diary examines the buying habits of people in the United States and the products and services that are bought by people in this country. Respondents to the CE Web Diary in the field complete the diary for two weeks for all of the items for which they spend money. In Round 1, three treatment groups were based on the way the expenditures were organized when presented to the participants. Group 1 had the expenditures organized by day and person, Group 2 had the items organized according to the type of purchase, and Group 3 had the items with no organization. We hypothesized that Group 2 would be the most efficient and satisfactory way to organize the expenses and that Group 3's lack of structure would make it the least popular. The results showed that the Group 2 was the most preferred organization style and Group 3 the least preferred, as predicted. Participants also had trouble with categorizing some items and over-reported alcohol purchases. In Round 2, participants were given "receipts" of the "purchases" to enter instead of narrative lists. Each participant was randomly assigned to one of three sorting conditions: Group 1 was instructed to sort the receipts in any way they liked, Group 2 was instructed to sort the receipts into the 4 CE Diary categories, and Group 3 was given no instructions. The results showed that many Group 1 and Group 3 participants ended up sorting receipts into the Group 2 categories

after learning of the diary format. Participants still had trouble categorizing some purchases and with over-reporting alcoholic beverages. There were some significant effects by age in both rounds that will be discussed.

16. Did the First Presidential Debate Really Matter? Evidence From the 2012 NORC Presidential Election Study

Rene Bautista, *NORC at the University of Chicago*; Tricia McCarthy, *NORC at the University of Chicago*; Kirk Wolter, *NORC at the University of Chicago*

One of the most salient events in the recent presidential election campaign, where the candidates discussed public policy issues, was the first presidential debate, held in Denver on October 3, 2012. Political and media analysts suggested that the candidates' performance in this debate had an effect on vote choice; however, little evidence was presented to support the argument. NORC at the University of Chicago conducted a pre- and post-election nationally representative survey of public opinion, with a focus on public policy issues, including the economy and the Affordable Care Act. The pre-election portion was fielded between September 24 and October 19. The fact that the debate took place in the middle of the fieldwork period represents a natural experiment that creates an opportunity to examine the impact of the debate on voter preference. Using multivariate regression techniques, this paper will use 'day of interview' as an instrumental variable to determine a potential effect of the debate, controlling for demographic characteristics, public policy knowledge on health issues, economic evaluation, and party identification. Additionally, NORC will collect actual vote choice among respondents who provided consent. Regression models will incorporate such data to the extent possible.

17. Social Network Analysis and Survey Response: How Facebook Data Can Supplement Survey Data

Adam Sage, *RTI International*

Social networking sites like Facebook and Twitter have resulted in the emergence of a type of data that is under-explored in the field of public opinion and survey research. Social network data is comprised of objects (typically people or groups) and the ties between the objects (e.g., relationships or transactions). Previously, obtaining these data to conduct thorough social network analysis was often impractically time consuming and costly. But increases in the ability to efficiently access such data have raised the potential for investigating new methods of analyses that may supplement current survey data, or otherwise fill holes in extant research where traditional analysis is limited. Addressing questions such as how objective measures of one's network differ from self-reported measures of such relationships, or how information flows and one's social context influence individual perception, thus survey results, are just a few examples of how survey and public opinion researchers might find value in social media and other Web 2.0 concepts. This paper demonstrates 1) how Facebook user data can be obtained through an application and utilized to reconstruct social networks, 2) how similar data scaled to a user's entire network can be analyzed to understand the formation of opinions, attitudes, and behaviors, and 3) how social network analysis of data native to social networking sites (e.g., a Facebook friendship) can enhance the interpretation and precision of such data when used to supplement survey data. Specifically, I describe an approach to developing a Facebook application that obtains friendship data from users, processes for obtaining a user's entire Facebook friendship network, and how I analyzed my personal social network on Facebook to produce measures. I then discuss how social network analysis techniques, such as cluster analysis and clique identification, can be used to supplement and provide precision to survey data.

18. Numbers, Numbers on the Dial, Which is the Fairest One on File? Cell or Landline? Home or Work? Findings from an ABS Longitudinal Study
Anna Fleeman, *Abt SRBI*; Tiffany Henderson, *Abt SRBI*; Patricia Vanderwolf, *Abt SRBI*; Kenneth J. Ruggiero, *Medical University of South Carolina*

Abt SRBI used an address-based-sampling (ABS) frame to select more than 200,000 addresses for a project fielding from 2011 through 2013. The use of an ABS was a result of landline RDD coverage issues and the need to target precise geographies. After addresses were selected, phone numbers were appended, if a match was available. Addresses able to be matched to a phone number were called, asked phone status (e.g., cell-phone-only or dual user), and screened for eligibility (12 to 17 in HH). Addresses unable to be matched to a phone number were sent a letter and questionnaire screening for presence of a 12- to 17-year-old and requesting contact information; two phone numbers were elicited. All available or provided phone numbers were dialed; if households were eligible, baseline phone interviews were conducted. Four and twelve months after completing the baseline, households were then re-contacted for follow-up phone interviews. If needed, all available numbers were called to maximize contact and response. We speculate that as cell-phone-only households are increasing, so too are those providing a “work phone,” which is typically a landline, on which to contact them. Presented findings will include the percentage of “work phones” provided, along with working number rates and response rates by matched phone status and type of phone –cell or landline, work or home – for all contacts during baseline and both follow-up interviews. As the reliance on ABS increases in the survey research field, knowing the best phone number on which to reach respondents is of critical importance. Further, the results provide insight as to the retention, contact, and response rates of all studies relying on an ABS frame.

19. Early Grade Reading Assessment – Using Tablet Technology and Efficient Survey Methodology in Developing Nations
Karol Krotki, *RTI International*; Michael Costello, *RTI International*

Implementing a standardized national survey in many countries poses challenges, not the least of which is how to incorporate national and sub-national cultural differences into the survey design framework. RTI International has developed a procedure for designing and implementing such a survey, EGRA/EGMA, Early Grade Reading/Mathematics Assessment to measure changes in education attainment across time and to compare countries and subpopulations within countries. The process also collects contextual demographic, socioeconomic, and education data to help in the analysis. In this presentation we describe how RTI has streamlined the sampling, data collection, data processing, and analysis to make each iteration efficient in its implementation and effective in producing the desired results. The data collection is carried out on tablets enabling standardized assessment, automated data correction, and speedy and error-free data transfer to a centralized server. This approach is now being widely used in many international education project evaluations and demonstrates how technology and good design can facilitate survey research in even very challenging circumstances.

20. Online Panels: Recruitment Based on “Hot Topics” – What are the Consequences?
Maria Andreasson, *University of Gothenburg*; Johan Martinsson, *University of Gothenburg*

Cost-efficient and representative recruitment to online panels is a persistent challenge for commercial enterprises and academic research alike. One method that is sometimes used is to take advantage of highlighting that the panel or the survey in question concerns a “hot topic” that most people are likely to find involving. This method can be exploited both with

probability based recruitments as well as with opt-in recruitments. This study compares the consequences of “hot topic”-recruitment both for opt-in recruitment and for probability based recruitment. During the summer 2012 four different surveys was fielded by the University of Gothenburg concerning a local “hot topic”: the introduction of congestion charges around the city of Gothenburg. In total four different surveys are compared: one from a pop-up ad on the major local daily newspaper website concerning the congestion charges, one survey to an opt-in sample from a general recruitment to the University of Gothenburg online panel, one probability sample from a postal invitation highlighting the issue of the congestion charges, and finally one probability sample from a general postal invitation to participate in an online panel. The outcomes that are compared include: recruitment rates, cost-efficiency, demographic and attitudinal representativeness. Special attention is paid to the hypothesis that “hot topic”-recruitment might help recruit those that are normally not interested in social or political issues, which might improve sample representativeness.

21. Relative Exposure: A Field Experiment Exploring the Influence of Public Opinion Polling Data on Voter Preference

Heather Knappen, Rochester Institute of Technology

This poster will present original research from a field experiment conducted on New York's 25th Congressional District race from July-August, 2012. A random sample of 200 registered voters was invited to participate in an experiment to determine whether exposure to public opinion polling data influences a voter's preference for the candidate leading in a poll. The experiment was conducted in three stages; first, each voter was called with a telephone poll to establish a baseline level of support for each congressional candidate. Next, half of the sample received one mailing and one robo-call with opinion polling data that showed one of the candidates clearly leading the other (59%-41%). A second telephone poll was used to determine whether voters who received the polling data were more inclined to support the candidate leading in the poll compared to voters who did not receive the polling data. The results from this experiment address critical questions about the influence of opinion polling data on voter preference. Although previous experiments have also addressed these questions, many of these studies have been concentrated in the laboratory. The benefit of conducting a field experiment is to provide a more realistic assessment of the influence of opinion polling data on voter preference within the context of a live campaign. Finally, this field experiment is one of the first to employ registration-based sampling for telephone survey research. The poster presentation will discuss several potential benefits of this sampling technique. Examples include improvements to polling analysis since voter files provide detailed demographic and voting histories, making it possible to more accurately identify “likely voters”. As the AAPOR community pursues a more sustainable future for public opinion research, this field experiment provides a good case study for the use of this sampling technique.

22. How Spending Money Can Save You Money: The Impact of Incentives on Speed of Response

Jennifer E. O'Brien, Westat

The effects of incentives on various aspects of survey administration continue to be active area of research. Features of the incentive such as the type (monetary vs. nonmonetary), amount (for monetary incentives), and the timing of the offer (prepayment vs. promised) have resulted in a few well-documented effects. Decades of research have demonstrated that, all else being equal, incentives increase participation rates and reduce refusal rates, cash incentives are more effective than non-cash incentives, prepayment of incentives is more effective than promised payment, the impact of incentives is greater in surveys with few, if any, other reasons to participate, and large incentives are not needed to recruit lower-

income respondents (Singer & Bossarte, 2006). Another, less well-documented, impact of the use of incentives concerns its influence on the speed of response. A handful of studies have observed that not only do incentives increase response rate, they also increase the speed of response (Czepiec, Landers, Hopkins, & Young, 1998; Gajraj, Faria, & Dickinson, 1990; Goldenberg, McGrath, & Tan, 2009; Hansen, 1980; Shettle & Mooney, 1999; Singer, Van Hoewyk, & Maher, 2000.) The results of the present study replicate this observation in a multi-mode, nationally representative household survey. In addition, we observed that the small monetary incentive used in this study not only resulted in faster submission of completed surveys but faster refusals as well. Thus, cases that were resolved quickly (whether they be completes or refusals) were removed from follow-up efforts resulting in savings to the study. In addition to presenting descriptive statistics, we will also present a cost savings analysis.

23. Well, Not Well, or Not Well at All? Evaluating American Community Survey (ACS) Data on School-Age Children Who Speak English With Difficulty
Angelina N. Kewal Ramani, *American Institutes for Research*; Amber Noel, *American Institutes for Research*

In 2010, approximately 22 percent of school-age children spoke a language other than English at home. As the U. S. population becomes more diverse, collecting accurate data on language use and language ability is increasingly important. Several studies report on the language proficiency of school-age children. The American Community Survey (ACS) reports on children who speak a language other than English at home and how well these children speak English. The U. S. Department of Education reports on the number of English Language Learners (ELL) in public elementary and secondary schools. These two sources reveal a different picture of school-age children with English language difficulties. Over the past five years, the percentage of school-age children who speak English with difficulty has remained steady (ACS), while the percentage of ELL students has consistently increased (Department of Education). This paper will examine these puzzling findings and evaluate the reliability of ACS language ability estimates for school-age children. The ACS includes a three-part question on language use and ability. Respondents who speak a language other than English at home are asked to assess how well they (or their children) speak English, either “very well,” “well,” “not well,” or “not at all.” Generally, respondents who report speaking English less than “very well” are considered to have some difficulty speaking English. The Department of Education’s Common Core of Data (CCD) identifies ELL students based on limited English language ability; therefore, these students speak English with some difficulty. In addition, the Department of Education Office for Civil Rights (OCR) collects more detailed information on ELL students including gender and race/ethnicity. ACS estimates for 2006 and 2011 will be compared with CCD and OCR data. Analyses will be conducted by gender and race/ethnicity. The research will reveal whether the ACS provides reliable estimates of language ability.

24. Page Reduction Experiment with Diverse Populations
Stephanie Lloyd, *Center for Survey Research, University of Massachusetts Boston*;
Carol Cosenza, *Center for Survey Research, University of Massachusetts Boston*;
Lee Hargraves, *Center for Survey Research, University of Massachusetts Boston*

Although it is increasingly common for health care organizations to survey their patients to assess the patient-centered care they provide, there is consistent pressure to minimize survey costs. Given the increasing printing and postage expenses associated with mailing paper questionnaires, one proposed way to lessen cost burdens is to minimize the number of pages in self-administered mail questionnaires, often by compressing text and formatting. Building on a previous experiment, the current project tested a CAHPS® Clinician & Group

(CG-CAHPS) questionnaire formatted to reduce its length from 12 to 4 pages to examine effects on data quality with different sample groups. The two groups in this experiment who were administered test questionnaires, included: (1) Spanish speakers, those sampled health plan members who requested Spanish materials, and (2) adults who were asked to respond about a sampled child. The survey, of which this study was a part, was funded by the Agency for Healthcare Research and Quality (AHRQ). All samples were drawn from a Medicaid population who were randomized to self-administer either the 12-page standard (Spanish and Child) or one of the test versions (n=500). Both the Spanish and Child test questionnaires were 4-page versions of the standard, using CAHPS guidelines, with the introduction and instructions at the top of the first page. A standard 3-contact mail administration protocol was followed. The current paper seeks to understand to what extent the cost savings associated with reducing the number of pages has an adverse effect on the quality of the resulting data. Response rates, item non response, substantive differences in answers between horizontal and vertical presentation of response alternatives, and mean CAHPS composite and rating measures will be compared across study arms. This study was in the field until September 2012, and the data will be analyzed by early 2013.

25. Putting a Little Religion Into Volunteer Activity

Robert K. Goidel, *Louisiana State University*; Belinda Davis, *Louisiana State University*

This paper began as a puzzle. Why did our state level estimates of volunteer activity in Louisiana differ so dramatically from CPS estimates? According to the CPS, only 1 in 5 Louisiana adults (20.9 percent) engage in volunteer activity. Our state level estimates from the 2012 Louisiana Volunteer Study (LVS) in contrast, place the number at just under half of all Louisiana adults. To understand the nature of these differences, we conducted a survey experiment in which respondents were asked either the CPS versions of the volunteer questions or the questions we have routinely asked. In the first part of the experiment, we tested the effect of including church as one of the organizations included in our standard volunteer question (listed below).

- VOL-1A: Have you done any volunteer activities in the last 12 months? I'm asking about activities for which you were not paid, except perhaps expenses, that you did in your neighborhood, or in other neighborhoods, at schools, churches, or for a volunteer organization.
- VOL-1B: Have you done any volunteer activities in the last 12 months? I'm asking about activities for which you were not paid, except perhaps expenses, that you did in your neighborhood, or in other neighborhoods, at schools, or for a volunteer organization.

In the second part of the experiment, we directly compared the CPS question wording to the LVS wording. The preliminary results indicate: 1) We estimate higher rates for volunteer activity even when we use the CPS question wording. This likely reflects other differences in terms of survey context, e.g., the LVS introduction cues respondents into the focus of the survey. It may also reflect differences in response rates or some combination nonresponse and subject matter; 2) Adding churches into the list of organizations in the LVS question significantly increases the volunteer rate.

26. First Contact Strategies for Web Surveys: Is a Phone Call or a Letter the More Effective Introduction?

Jill Connelly, *NORC at the University of Chicago*; Micah Sjoblom, *NORC at the University of Chicago*; A. Rupa Datta, *NORC at the University of Chicago*; Peter Hepburn, *NORC at the University of Chicago*

The objective of the National Survey of Early Care and Education (NSECE) is to document the nation's current use and availability of early care and education, and to deepen our understanding of the extent to which families' needs and preferences coordinate well with providers' offerings and constraints. The NSECE included a survey of home-based child care providers who were licensed or otherwise registered with state agencies. The survey included Web data collection, with phone or in-person follow up as needed. Individuals who provide care to children in a home-based setting tend to be older or lower-income or in other demographic subgroups that have lower Internet usage rates. In order to encourage participation by Web, a \$35 gift card was offered to complete the interview online. We had phone numbers, but no mailing or e-mail addresses for sampled individuals. We designed an experiment with 1,300 providers to test whether it would be more efficient to 1) send a letter or e-mail as a first contact based on locating efforts that didn't involve personal contact with the respondent, or 2) make a gaining cooperation phone call first, to introduce the study and then request mailing or e-mail information to send the Web survey request. Our evaluation includes comparisons of effort required, success rates in reaching respondents through initial contact attempts, cooperation with the initial request, and final cooperation rates.

27. How Did the 2012 U.S. Presidential Campaign Season Affect Media Consumption and Behavior?

Daniel Hutchison, *Arbitron, Inc.*

U.S. Presidential Election campaign seasons have several key media events including both Television and Radio coverage of the National Conventions and the Presidential debates. News broadcasts during the weeks leading up to the election carry coverage of campaign efforts and culminate with Election Day coverage. Additionally, the overall environment of the 2012 campaign season was shaped by a high utilization of political advertising, often negative, throughout the nation. Specific states and markets received varying levels of this advertising. Specifically, metros including districts with races for seats in the House of Representatives and the Senate, perceived to be of "High-Value" by the national political parties, attracted a higher level of advertising support from the parties themselves and from independent Political Advocacy Coalitions as well. This led to an intensity of campaign-related advertising that many perceived to be excessive. Arbitron PPM is a system to passively collect Radio and Television media use over time among an on-going panel of respondents. This system replaced the traditional paper Radio and Television self-report diaries previously used in 47 top U.S. metros. This paper will explore the result of specific media events, news broadcasts and, to a lesser degree, the result of political advertising on Radio listening and Television viewing behavior. Results will include comparisons of overall listening and viewing before and following the campaign season across the 47 metros measured by Arbitron's PPM system. Analyses of listening and viewing will be presented by radio station format and for stations carrying the special media events overall and by age, gender, and racial groups. Reviewing these results will add to our knowledge of the potential impact of the media on public opinion during this election season.

28. Crowd Coding: Increasing the Time and Cost Efficiency of Common Research Tasks
Michael Jugovich, *NORC at the University of Chicago*; Patrick Van Kessel, *NORC at the University of Chicago*

In recent years, many online crowdsourcing platforms have been developed and now provide organizations with the opportunity to outsource simple yet labor-intensive tasks to a large pool of individuals around the world. With the advent of services such as Amazon Mechanical Turk, which offers the ability to easily create Human Intelligence Tasks for distribution across a user base active both during and after normal business hours, researchers now have the ability to leverage crowdsourcing technology to alleviate some of the costs associated with straightforward coding tasks traditionally allocated to in-house resources. Seizing this opportunity, NORC has developed a “crowd coding” software package that allows researchers to quickly deploy custom assignments to Mechanical Turk, with applications for research projects involving not only traditional designed data but organic data as well. Examples include sentiment and relevancy analysis of social media data, and the rapid and inexpensive construction of context-specific training datasets for machine learning algorithms to be deployed on Big Data collections. This presentation will focus on a series of case studies that explore the effectiveness of crowd coding compared to traditional manual coding, measured across three dimensions: time, cost, and accuracy. It will conclude with a discussion of the pros, cons, and potential future applications of the technology.

29. Use of Paradata to Predict Participation in a Randomized Control Trial Intervention
Harmoni Noel, *American Institutes for Research*; Simone Robers, *American Institutes for Research*; Grace Wang, *American Institutes for Research*; Alex Ortiz, *American Institutes for Research*; Amy Windham, *American Institutes for Research*; Steven Garfinkel, *American Institutes for Research*; Kristin Carmen, *American Institutes for Research*

Paradata is increasingly being used to monitor response rates, conduct respondent validation, evaluate interviewer performance, and determine cost efficiencies in the survey administration process (Kreuter, Couper, & Lyberg, 2010). Paradata has also been used in an adaptive design framework to tailor interventions to a subgroup of the sample to achieve higher response rates (Couper & Wagner, 2011). This paper uses data from a randomized control trial study with a pre/post survey intervention design to examine the use of paradata such as pre-survey completion time as an indicator of likelihood to participate in the intervention. The authors hypothesize that participants with shorter response times are less likely to attend. If this hypothesis is supported, it suggests that survey completion could be used in multi-stage research to tailor follow-up strategies to increase participation in subsequent stages. In this study 1,747 participants were randomized into four experimental conditions or the control group across four locations. The four experimental conditions represent four methods for conducting public deliberations. These deliberations are designed to obtain informed perspectives on complex topics similar to those that arise frequently with respect to healthcare and health research decision making. The four methods have varying levels of respondent burden, vary between in-person and online formats and have varying attendance rates. Participants were recruited into the study before pre-survey administration which led to a high overall response rate of 94%. First, the authors will conduct a non-response bias analysis to compare intervention response rates by respondent characteristics (race/ethnicity, gender, age, occupation, and level of education), recruitment location, and experimental method to see if response propensity varies by these subgroups. Second, the authors will examine whether pre-survey completion time is related

to subsequent participation in the intervention and whether different variables interact with the potential effect of completion time.

30. Designing Questions to Measure Number of Sex Partners Among At-Risk Youths in ACASI (Audio Computer-Assisted Self-Interviewing)

Kerryann DiLoreto, *University of Wisconsin Survey Center*; Jennifer Dykema, *University of Wisconsin Survey Center*; Jessica Price, *University of Wisconsin Survey Center*; Nora Cate Schaeffer, *University of Wisconsin Survey Center*

A central concern for questionnaire designers is how to design questions to accurately measure the frequency of sensitive behaviors. For interviewer-administered surveys, past research indicates that higher reports of sensitive behaviors may be obtained using: open (versus closed) questions (e.g., Blair et al. 1977) and audio computer-assisted self-interviewing (ACASI) (versus interviewer administration) (e.g., Turner et al. 1998). Little research, however, explores how differences in question wording affect responses using ACASI. We implemented an experiment using ACASI in which respondents reported the total number of sex partners they had in their lifetimes and the last year using one of three question formats: (1) closed-low frequency (from Add Health) that used the categories 0, 1-2, 3-4, or 5 or more partners; (2) closed-high frequency that used the categories 0, 1, 2, 3-4, 5-6, 7-8, 9-10, or 11 or more partners; and (3) open-total frequency that allowed respondents to enter a value for the total number of partners. Data are provided by the Midwest Young Adult Study, a longitudinal in-person study of young adults transitioning out of foster care. Current data are from Wave 5 (2010-2011) in which 82% of the baseline sample (n = 590) were re-interviewed. This hard-to-reach population is characterized by high engagement in behaviors with negative consequences and low literacy levels. While we find no differences among the question formats in reporting about sex partners in the past year, the open-total format is associated with lower reporting among men and higher reporting among women for lifetime partners, and less missing data than the closed formats. These results are consistent with research assessing the quality of reporting about sexual partners which finds that men overreport and women underreport sex partners (Laumann et al. 1991), and add to the body of research that recommends using open questions to measure sensitive behaviors.

31. Household Composition and Child Wellbeing: Using Quantitative Data to Construct Narratives to Inform a Research Agenda

Catherine C. Haggerty, *NORC at the University of Chicago*; Kate Bachtell, *NORC at the University of Chicago*; Nola duToit, *NORC at the University of Chicago*; Ned English, *NORC at the University of Chicago*

Due to the deinstitutionalization of marriage, high levels of divorce, and an increased acceptance of cohabitation and single parenthood, there is a changing array of families in American households (Stacey 1996, Thistle 2006). Du Toit et al. (2011) used data from two waves of the Making Connections Survey, a study of disadvantaged urban communities, to examine different types of households, the extent of change in household composition, and differences in the effect of various household structures on a variety of economic measures of child wellbeing. They observed large proportions of households that do not fit the traditional nuclear family model and are not accounted for in conventional family studies. These non-traditional households differ along several measures of economic wellbeing. Changes in the composition of these different households further impact their economic stability and, therefore, child wellbeing. Building on this quantitative research and using quantitative data, we used a grounded qualitative approach to develop case studies of four types of households: two parent, single parent, non-parent and extended family households to further explore the characteristics of distinct household types, how they changed over

time, and how their unique qualities impacted child wellbeing. This methods brief presents the process of developing these case studies to further explore the characteristics of these distinct household types which informed the next steps in our research agenda.

32. Over sampling Young Adults on Cell Phones

Randal ZuWallack, *Abt SRBI*; Thomas Duffy, *RTI International*; Matthew Denker, *Abt SRBI*

Young adults are often a key research group in public health and public safety surveys. Many research organizations, such as the National Highway Traffic Safety Administration, conduct surveys with oversamples of this age cohort to ensure sufficient data to analyze driving behaviors and attitudes. In a recent national survey, nearly 40% of the cell phone interviews were with respondents under age 35; the same survey yielded young adults less than 10% of the time on landlines. It is clear that cell phones are an efficient method for increasing the sample size for young adults. We conducted a cost-benefit analysis to determine the best sampling design when young adults are a subpopulation of interest. Optimal allocations that account for landline and cell cost differentials are not optimal for reaching this population because the costs will favor the landline sample, resulting in an undersample of young adults. We compare costs and benefits for three dual-frame designs, 1) based on the overall optimal allocation, 2) based on a screening oversample of young adults, 3) and one with a higher allocation to cell phones. All designs are based on a fixed cost and compared on the overall sample size, the sample size of young adults, and the resulting design effects.

33. Those are the Breaks: Incumbents, Challengers and the Distribution of Unallocated Votes in Pre-Election Polls

Christopher P. Borick, *Muhlenberg College Institute of Public Opinion*; David G. Wegge, *St. Norbert College*

In almost every case pre-election polls contain a portion of voters who identify themselves as “undecided” in terms of their candidate preferences in an upcoming election. In 2012 for example, about 7% of pre-election poll respondents in Senate and gubernatorial election polls conducted in the week before the election identified themselves as undecided in terms of their voting plans. Of course on Election Day those undecided voters either select a candidate or decide not to vote at all, leaving no voters unallocated in the final results. So how do the undecided voters in pre-election polls break in terms of their ultimate decision? For many years there was evidence that most unallocated votes broke towards challengers in statewide races. However in the last decade it appears that there has been an increasing share of unallocated votes being captured by incumbents seeking reelection. In this paper we examine the unallocated vote in the 2012 election and the role that incumbency, party affiliation and other candidate characteristics played in terms of the distribution of the unallocated voters in the final results of Senate, House and gubernatorial races.

34. God, Money, Politics & Science: The Role of Religion, Conservative Economic and Liberal Social Attitudes on Perception of Science in the Last Weeks of the 2012 U.S. Presidential Election

Kristin Runge, *University of Wisconsin – Madison*

This study uses a two-wave panel design to examine the effects of perceptual filters in predicting science-related opinion and media use during the weeks immediately prior to and after the 2012 U.S. Presidential election. The first wave was conducted in the two weeks prior to the first candidate debates (September 25, 2012 through October 8, 2012), and the second wave was conducted after the election (field dates November 14, 2012 - November 21, 2012). A total of 1,401 respondents were segmented into 4 attitudinal clusters based on

religiosity, liberal/conservative economic attitudes and liberal/conservative social attitudes. In our preliminary analysis of the first panel wave, we find that respondents clustered into one of four segments: 1) high religiosity with conservative ideologies, 2) high religiosity with liberal or moderate ideologies, 3) low religiosity with conservative ideologies, and 4) low religiosity with liberal or moderate ideologies. Initial analysis indicates that response to 'How much guidance does religion provide in your everyday life?' is the strongest determinant of cluster membership among the attitudinal bases variables. After controlling for demographic characteristics, cluster membership predicts a number of items including likelihood of voting for President Obama or Governor Romney, media habits, support for federal funding of science, support for free market regulation of nanotechnology, benefit perception of nanotechnology, synthetic biology and stem cell research, as well as trust in university scientists, corporations, environmental organizations and religious institutions. Final analysis will show how panel members voted and determine if attitudes and behaviors changed during the final weeks of the election. Implications of results for media effects, science communication and political communication research will be discussed.

35. Public Sentiments Online: New Tools of Measurement Combining Human- and Computer-Based Coding

Leona Yi-Fan Su, *University of Wisconsin – Madison*; Xuan Liang, *University of Wisconsin – Madison*; Nan Li, *University of Wisconsin – Madison*; Dietram A. Scheufele, *University of Wisconsin – Madison*; Dominique Brossard, *University of Wisconsin – Madison*; Michael Xenos, *University of Wisconsin – Madison*

The Internet provides researchers with a wide variety of tools for tapping opinion expressions on Web-based platforms. This study uses a new content analysis method for tapping opinion expressions in online Big Data environments. Based on a carefully constructed keyword search about scientific topics, a series of Twitter posts are first randomly pulled from publicly-available Twitter accounts. The selected content is interpreted and analyzed by trained coders and then translated into appropriate categories. Computational software (Crimson Hexagon) then extracts the linguistic patterns from the coded examples and uses the resulting algorithms to track these patterns in every captured tweet. In our method, human coders no longer serve as text-level analysts; instead, we capitalize on human coders to translate extract sentiments and latent meanings from the Tweets (equivalent to building a codebook in traditional content analysis) and use the resulting algorithms to guide the computer-based analysis. In other words, computer algorithms inductively determine the patterns of underlying content identified by human coders, and then apply the learned patterns for large scale data processing. Our study also provides empirical verification that this method can accurately analyze defined communication content, sentiment and topics from large-scale datasets. Using nuclear energy as one exemplar, we track public opinion expressed on Twitter before and after the Fukushima Daiichi disaster in order to examine if variations in our sentiment coding reflect changes consistent with these external influences. We also compare nuclear energy to other scientific issues to demonstrate how our method proves to accurately track public sentiments across issues without introducing false positives or other biases. Our results suggest that this method works well in capitalizing on the strengths of human coding in terms of preserving sentiment validity while relying on computer-based coding to reliably process large-scale data of online opinion expression (e.g., millions of Twitter posts).

36. Turnout Validation of Survey Respondents in New Jersey

Ryan Tully, *Princeton University*; Amy Lerman, *Princeton University*

It is commonly observed that self-reported voter turnout in surveys is substantially higher than actual turnout. In previous studies, researchers have attempted to use government

records to validate self-reported voter turnout among individual survey respondents. More recently, Berent, Krosnick, and Lupia (2011) attempted to validate self-reported voter turnout among participants in the 2008 American National Election Study (ANES) using government records. Their study found that the “success” of turnout exercises in previous studies may be due to an inherent bias that “people who choose to participate in surveys also choose to participate in elections at a higher rate than people who do not participate in surveys” (p. 8). Our study expands on this initial finding by conducting a turnout validation exercise for a series of surveys conducted in central New Jersey in 2011. In our analysis, we compared respondents based on various aspects of survey participation, including those who volunteered to participate in an online panel or volunteered personal information with those who do not. Overall, we found that respondents who opted into the online panel or volunteered personal information were significantly more likely to accurately report voter turnout than those who do not. Furthermore, we also found that various demographic characteristics, such as age, race, educational attainment, and income, correlated with significant differences in the accuracy of self-reported voter turnout among survey respondents.

37. Who is Really Ahead in Election Polls? Practical Guidance on Assessing the Gap Between Two Candidates

Kien Le, *Social and Economic Survey Research Institute, Qatar University*; Abdoulaye Diop, *Social and Economic Survey Research Institute, Qatar University*; Darwish Alemadi, *Social and Economic Survey Research Institute, Qatar University*

In election poll results, the proportions favoring candidates and the survey sampling error are usually reported. However, it is hard to assess if the gap between any two candidates is statistically significant or not based on this information. This note provides an alternative measurement of sampling error for this assessment purpose. We detail the calculation steps in STATA and SPSS programs to handle polls based on simple random sampling and also polls based on more complicated designs.

38. Are Declining Response Rates Only a Symptom of a Bigger Problem?: Assessing Trends in Survey Response Quality Between 2005 and 2013

Curtiss Cobb, *GfK Knowledge Networks*

In May 2012, Pew Research shocked the research community by publicly stating an already widely known fact—there has been a general decline in response rates that is evident across nearly all types of surveys. Pew offered as an example that its typical telephone survey response rate in 1997 was 36% and is just 9% today. At the same time, greater effort and expense are required to achieve even the diminished response rates of today. These challenges have led many within and outside the survey research community to question whether surveys are still providing accurate information. Non-response is only one way that the quality of surveys may have changed over time. Unit non-response may merely be a harbinger of the declining quality of responses even among those that do respond.

Satisficing behavior and item non-response may be increasing over time as well. Or, it is also possible that non-responders were the “bad” respondents in previous years, leaving only those that optimize their responses and measurable satisficing behavior has decreased. Of course, measurement error and non-response may be completely unrelated. This study seeks to examine trends in data quality to determine whether response quality is also changing over time. We use 7 years of data profile data from GfK’s probability-based Internet panel, KnowledgePanel®. We test whether item non-response, breakoffs, straight-lining, speeding, and other satisficing and sub-optimal response behavior is increasing, decreasing or remaining constant over time. We explore these trends in general and within demographic groups.

39. Measuring Parental Engagement With Children's Schools

Beth Schueler, *Harvard Graduate School of Education*

Researchers have repeatedly demonstrated that parental engagement with schools is associated with positive educational and social outcomes for children (Walker, Wilkins, Dallaire, Sandler, & Hoover-Dempsey, 2005). However, to accurately measure parent engagement, new tools are needed that take advantage of best practices in survey design. This poster outlines the development of a survey scale to assess parent perceptions of their engagement with their child's school. We employed Gehlbach and Brinkworth's (2011) 6-step process for scale development that front-loads input from academics and potential respondents during item-development to establish evidence for validity with regard to both populations. First, we conducted a literature review to define the construct and identify potential indicators. Second, we conducted open-ended interviews with diverse groups of parents to learn how they conceptualized of school engagement. Third, we systematically compared literature review and interview results, noting distinctions in the language academics and parents used. These findings informed our item phrasing. Fourth, we crafted preliminary items that reflected key factors of engagement. Fifth, we subjected our items to an expert review process regarding the relevance, comprehensiveness, developmental appropriateness, and cultural appropriateness of our items (Rubio, Berg-Weger, Tebb, Lee, & Rauch, 2003). Sixth, to ensure that parents understood our items as intended, we conducted a "cognitive pretesting" procedure. We asked parents to rephrase the questions in their own words and think aloud when answering the questions and then edited some items for clarity. Finally, we conducted three studies with large national samples of parents (n=385; n=253; n=531) to gather evidence of reliability and convergent/discriminant validity. Through confirmatory factor analysis we identified a theoretically grounded factor structure that fit the data well. The poster will describe the implications of our process for scale validity and ways researchers and Pre-K – 12 schools can use the scale to aid school improvement efforts.

40. The Case for Town Hall Debates: The Effects of the Press and Public Agendas on Voter Acquisition of Campaign Knowledge

Jason Turcotte, *Louisiana State University*

An uninformed and unmotivated electorate has plagued American democracy for decades. Americans know very little about their public officials and their stances on issues. With the media devoting more attention to negativity, tactics, attack ads and horserace coverage, voters have fewer avenues for learning about the candidates and substantive issues. Political debates are one of those avenues, and perhaps serve as the only remaining campaign event maintaining a mass audience. Using the 2008 National Annenberg Election Survey data, this paper explores the relationship between exposure to the 2008 U. S. presidential debates and political knowledge. More specifically, this project explains whether survey data can reveal a more nuanced understanding of debate effects by extending previous scholarship to account for effects differences among formats. I hypothesize that exposure to all three general election debates holds a positive relationship with political knowledge but, also, that the town hall debate – debates in which the electorate has a hand in shaping the debate agenda – foster greater knowledge gains than traditional media moderated debate formats where the press serves as sole gatekeeper of the discourse. After controlling for a number of other variables known to influence political learning and political knowledge, I find support for both hypotheses. As criticisms of debate formats and moderating grew even louder in the 2012 U. S. presidential debates, these findings hold numerous implications for democratic process and offer some preliminary evidence that more participatory debate formats may improve political knowledge.

41. Blogging Nanotechnology: Public Discourse Around Emerging Technologies in the Blogosphere

Xuan Liang, *University of Wisconsin – Madison*

Communication environments in this information age are experiencing rapid changes and the Internet emerges as one of the dominant channels for science information (“Science and Engineering Indicators,” 2012). New media forms, such as blogs, forums and podcasts, can serve as public spaces for audiences to share knowledge, develop ideas about science, and interact with scientists in a timely fashion (Birch & Weitkamp 2010). This raises important questions about the types of user-generated information and opinions surrounding emerging technologies, such as nanotechnology, that audiences may encounter in blogs. In order to explore the landscape of blog traffic about nanotechnology, we use a computational linguistic software to analyze a census of all English-language nanotechnology-related blog posts generated between January 1, 2009, and October 31, 2012. Results of content analysis and sentiment analysis on a total number of 680,790 related posts show that nanotechnology is depicted in a comprehensive and comparatively positive picture in the blogosphere. Overall, most of the blog posts presented information about nanotech related consumer products, followed by discussion about business, national security, medicine, EHS (Environmental Health and Safety), basic research and energy. Thirty-six percent of blog posts expressed optimistic opinions, 32% expressed neutral opinions and 32% expressed pessimistic opinions. Interestingly, we found that scientists’ latest research was reflected in the perceivable fluctuations of some topics covered in the blog posts. Our results have significant implications for the understanding of the open discourse of nanotechnology in the blogosphere, and more importantly, how new media on the Internet reflects and shapes public opinion of this emerging technology.

42. Is Deliberative Science Possible? Examining the Links Between Informational Factors, Scientific Knowledge, and Attitude Extremity

Nan Li, *University of Wisconsin – Madison*; Dominique Brossard, *University of Wisconsin – Madison*

In the past decades, U.S. citizens have increasingly been asked to engage in the decision-making process related to science policy with a high level of public interests at stake. Those who hold strong opinions about the issues at hand are more likely to participate in public discussions and to express themselves openly. Studies have shown that the strength of individual attitudes can be influenced by a variety of factors, including the heterogeneity of networks and the nature of the information environment one is constantly exposed to. In this study, we examine whether and how people’s attention to news and entertaining contents on mass media may influence the extremity of their attitudes toward the issue of nuclear power. In addition, we test whether interactive online communication makes people develop strong opinions on this issue. Using data from a nationwide online survey carried out in 2010 (N = 1,138), this study finds that higher levels of attention to news in newspapers and television, as well as more frequent interpersonal talk, make people develop more extreme attitudes toward nuclear power. The relationships between media use, interpersonal talk, and attitude extremity, are mediated by the level of factual knowledge about this issue. In contrast, interactive online communication is not significantly related to attitude extremity. Results hence suggest an absence of the so-called “echo-chamber” effects of the Internet regarding controversial scientific issues. In fact, people tend to develop extreme attitudes toward nuclear power based on the knowledge that is obtained either from interpersonal talk or newspaper and television news.

Friday, May 17
3:15 p.m. – 4:15 p.m.
AAPOR Demonstration Session #2

Mathematica's Survey E-Tool: Assisting Third-Party Data Collection
Kristina P. Rall, *Mathematica Policy Research*

With the rapid advancement of data collection technology, including Web-based instruments and handheld devices, it's easy to lose sight of the continuing need by some organizations to collect data the "old-fashioned" way, using paper questionnaires or basic Windows applications. A growing number of such organizations are seeking technical assistance to conduct their own surveys rather than contracting them out. However, they may lack funding for devices such as laptops, tablets, and smartphones and may not have Internet access in locations where the survey is conducted, limiting their ability to use the most modern survey modes. The Centers for Medicare & Medicaid Services (CMS) faced this constraint in administering its Money Follows the Person (MFP) project. MFP provides demonstration grants to 44 states to help them reform their financing and service designs for long-term health care, the ultimate goal being to measure the costs and benefits of transitioning some Medicare patients from institutional to community care settings. The Quality of Life (QoL) survey is currently being conducted to evaluate MFP; in administering the QoL, participating states follow patients from one care setting to another, surveying them at several times in multiple locations. To help states gather and submit data for QoL, Mathematica Policy Research developed the "Survey E-Tool." This user-friendly database streamlines the process of manually entering data collected on hardcopy forms and enables states to transmit data through Gentran when an Internet connection is available. The tool is programmed to account for different versions of Access used by state offices. Without this standard system, data would be submitted as Excel files with no consistent layout, necessitating additional time and expense merging files to conduct analysis.

Colectica for Microsoft Excel: Increasing Transparency Using Open Standards
Dan Smith, *Colectica*

Colectica is a suite of modern metadata management software that is used to document public opinion and survey research methodologies and data. This demonstration will introduce the new Colectica for Microsoft Excel software, a free tool to document statistical data using open standards. There is often inadequate transparency of research methods when results of opinion polls and behavioral science research are disseminated. Colectica allows organizations to increase their openness and credibility through standardized documentation of their data collection, research process and resulting data. The software implements leading open standards including the Data Documentation Initiative (DDI) Lifecycle version 3 and ISO 11179. Using this software allows survey organizations to both better educate survey sponsors and the public on their methodology and increases the organization's reputation for performing credible scientific research. The free Colectica for Excel tool allows researchers to document their data directly in Microsoft Excel. Variables, Code Lists, and the datasets can be globally identified and described in a standard format. Data can also be directly imported and documented from SPSS and Stata files. The standardized metadata is stored within the Excel files so it will be available to anyone receiving the documented dataset. Code books can also be customized and generated by the tool, and output in PDF, Word, Html, and XSL-FO formats.

Roper Center: Archiving Services and Access Tools
Lois Timms-Ferrara, Roper Center for Public Opinion Research
Marc Maynard, 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. Preserving data for long term access requires vigilant review of all data and documentation, standardization of data formats, and ongoing attentiveness to aging and new technologies impacting those data formats. This year, new procedures have been adopted to clearly identify specific features of the data coming into the Center that mirror those of AAPOR's Transparency Initiative. By more completely documenting the details of publicly released survey data at the point of ingest, the objectives of the TI and the Center to better inform poll consumers may be achieved. Come and see how this new process works. This spring the Roper Center released a set of enhanced services impacting access to some 20,000 U.S. and international survey datasets archived at the Center, as well as, iPOLL a database of more than 600,000 U.S. 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. Bring your research questions to this session!

Friday, May 17
4:15 p.m. – 5:45 p.m.
AAPOR Concurrent Session F

**Questionnaire Design and Data
Quality**

**Associations Between Interactional Indicators of Problematic Questions and
Systems for Coding Question Characteristics**

**Jennifer Dykema, University of Wisconsin Survey Center; Nora C. Schaeffer, University
of Wisconsin Survey Center; Dana Garbarski, Center for Women's Health and Health
Disparities Research**

Writing survey questions requires attention to the conceptual and operational definitions of survey concepts as well as to the technical issues that arise in composing items. These technical issues are examined in a body of research that considers how characteristics of questions (e.g., the number of categories to include in a rating scale) affect responses, their distributions and associations with other variables, and their validity and reliability. While the analysis of the properties of questions has led to the development of several ad hoc and formal systems for coding characteristics (e.g., Problem Classification Coding System (CCS) (Forsyth et al. 2004), Question Appraisal System (Willis 2005), and Question Understanding Aid (QUAID) (Graesser et al.), these systems vary considerably in the assumptions that underlie which

characteristics they identify as problematic, which characteristics are compared, and how dependencies among characteristics are taken into account when writing questions. Our paper has several goals. First we review and synthesize the literature on question characteristics and the systems for coding characteristics. Second, we analyze the administration of questions about physical and mental health from 350 digitally recorded and transcribed interviews with older adults in the Wisconsin Longitudinal Study. Interviewer-respondent interaction has been coded in Sequence Viewer and we have also coded the questions' characteristics using several different coding schemes. We identify interactional behaviors that have been associated with poorer data quality and use multi-level models to determine which coding systems are best at predicting problematic outcomes, including interviewers misreading questions and respondents expressing uncertainty and requesting clarification. Our analysis adds to the small but growing body of research concerning the effects of question characteristics on interaction and data quality. Our results have implications for designing questions and interviewing procedures with an emphasis on health surveys of older adults.

Interaction Between Questionnaire Design and Interviewer Performance

Pat D. Brick, *Westat*; Catherine Billington, *Westat*; Sarah Dipko, *Westat*; J. Michael Brick, *Westat*

There is a large literature devoted to structuring and controlling the behavior of the interviewer in telephone and in-person surveys with the goal of improving the quality of the data collected. This literature discusses such topics as interviewer error and interviewer effects. The behavior of the interviewer is the focus of these studies and interviewer behavior is treated as an exogenous variable. The interviewer effects are often described as increasing the variance of the estimate rather than causing biases because, in the models, interviewer effects are assumed to have an expected value of zero across interviews (O'Muircheartaigh and Campanelli 1998). These studies are helpful, but treat interviewers in isolation for other features of the survey. We suggest that this approach is incomplete because in many cases, the behavior of the interviewer is a function of characteristics of the questions being administered. We suggest that some survey questions may generate greater interviewer effects than others due to the way survey questions are constructed. Our research links the questionnaire design characteristics to the interviewer effects. We begin our investigation by conducting an expert review on questionnaire items in a CATI survey. Items are classified as either having potential problems or being well-constructed. As a complement to the expert review, we examine and tabulate the comments entered during the interview for all items. Our analysis assesses whether problematic questions generate more comment entries than well-designed questions. The second part of the analysis deals with interviewer effects linked to the questionnaire design characteristics. The goal is to determine which questionnaire items experienced greater and lesser interviewer effects. Ultimately, we seek to evaluate the hypothesis that interviewer effects are at least to some extent a function of questionnaire design characteristics and that crafting high quality survey questions is the best way to control interviewer behavior.

An Examination of the Relationship Between Pretest Method Results and Data Quality

Aaron Maitland, *Westat*

Many research studies collect data through survey questionnaires. In order to enhance the validity of the findings from these studies, it is important for the studies to employ questions that

minimize measurement error. A diverse range of question evaluation methods are available for detecting measurement error in survey questions. Ex-ante question evaluation methods are relatively inexpensive, because they do not require any data collection from actual survey respondents. Other methods require data collection from respondents either in the laboratory or in the field setting. A major gap in the literature is the general lack of evidence that the problems identified by these methods are actually problems as assessed by traditional quality standards such as reliability or validity. Although one would expect these methods to identify questions that produce low quality data, behavior coding is the only technique in the literature that has been consistently shown to predict the reliability and validity survey questions (Dykema, Lepkowski, and Blixt 1997; Hess, Singer, and Bushery 1999). This paper addresses the important gap in the literature about whether the problems identified by question evaluation methods lead to lower quality data. The research in this paper investigates how effectively these methods predict the reliability of survey questions as measured by test-retest correlations obtained from repeated measurements of sample respondents. The study uses question evaluation results from a few ex-ante methods such as expert review and QAID, laboratory methods such as cognitive interviewing, and field methods such as behavior coding and response latency to predict the reliability of survey questions. In addition, the study evaluates how the results from question evaluation methods relate to other data quality indicators such as item missing data.

Can Google Consumer Surveys Help Pre-Test Alternative Versions of a Survey Question?: A Comparison of Results from Cognitive Interviews and Google Consumer Surveys on Alternate Forms of Two Questions

Michael Stern, *NORC at the University of Chicago*; Vincent Welch, *NORC at the University of Chicago*

During the 1990s cognitive interviewing in its various incarnations (e.g., concurrent think-aloud, retrospective think-aloud, focus group discussions, probes, and memory cues) became the primary means for evaluating questions (see Lessler and Forsyth, 1995; Conrad and Blair, 1996; Willis and Schechter, 1997; Tourangeau, Ripps, and Rasinski, 2000). By examining the cognitive processes respondents went through while interacting with a survey question, survey methodologists uncovered how small manipulations in the wording of questions influenced respondents' answers. Another way researchers have historically uncovered the effects of question wording is through experimental field tests where several versions of a question are randomly assigned to a subsample of respondents. Over the past decade, researchers have embedded the bulk of these experiments in Web surveys among undergraduate students due to the affordability of implementing experimental designs in this mode and the technological acuity of college students. Still, if a researcher wanted to assess a single-item among a large heterogeneous audience, their options were limited. Google non-probability Consumer Surveys may provide a solution. However, two questions remain to be answered. First, how do results from these non-probability surveys compare to those from proven cognitive interview techniques. Relatedly, what is the value added by conducting such experiments. In this paper, we compare results from alternate forms of two questions that were tested at NORC at the University of Chicago with 2-waves of cognitive interviews and with Google Consumer surveys (N=4,000) to answer these questions. The results suggest that the Google Consumer survey data do complement the findings from cognitive interviews and the inclusion of the inferred weighted demographics data are useful for use in certain types of studies.

An Empirical Test of the Effectiveness of Cognitive Testing in Improving Question Wording

Martha Stapleton, Westat; Jeffrey Kerwin, Westat; Jennifer Crafts, Westat; Jasmine Folz, Westat

Cognitive interviewing has become an accepted survey industry best practice due to the face validity of basing question revisions on feedback elicited from representatives of the survey target population. Despite expectations that such revisions improve data quality and reduce burden, little empirical evidence supports the effectiveness of cognitive interviewing (Willis, 2000; Willis and Schechter, 1997; Forsyth, et al., 2004). Our experiment focused on questions with comprehension problems. We provide evidence that issues revealed and addressed through cognitive interviewing are associated with improved survey outcomes, including data quality and response burden. We conducted a between-subjects experiment with approximately 20 items organized as: 1) control set (“original” questions -- before cognitive testing), and 2) experimental set (same questions modified on the basis of cognitive test results). CATI interviews were administered to a sample of 200 U.S. general population, English-speaking adults. Through random assignment, each respondent received a different mix of both control and experimental questions. At the interview end, respondents were asked to explain the meaning of two questions in their own words so that we could judge whether their responses to those questions were “accurate” (consistent with the question intent). Interviews were timed and recorded for later behavior coding. Participants received a \$10 incentive. We examined missing data to evaluate whether the cognitively tested questions resulted in fewer “don’t know” / “can’t answer” responses, compared to the “original” questions. We compared time to respond to evaluate whether the cognitively tested questions required a lower average time per recorded answer. We compared the follow-up probes to the survey questions to evaluate whether responses to the cognitively tested questions appeared more accurate than responses to the original questions. Future behavior coding analysis will compare control versus experimental groups on requests for repeats and clarifications and the match between respondent answers and response categories.

Methodological Briefs: Combating Nonresponse

The Impact of Incentives in a National RDD Survey **Kelly Daley, Abt SRBI**

While there is considerable empirical support for pre-paid monetary incentives (Church 1992; Singer et al. 2000), the benefits from post-paid incentives – particularly in RDD surveys – are less clear (Singer et al. 2000; Gelman et al. 2003). Designing an effective post-paid incentive is particularly challenging when the survey features both a screener and an extended interview. Prior research suggests that incentives offered for the extended interview may be more cost-effective than incentives offered for the screener (Arbitron 2003; Cantor et al. 1998; Kropf et al. 2000; Singer et al. 2000), but gaining participation at the initial stage is often the most challenging component of RDD surveys. The 2012 Family and Medical Leave Survey of Employees conducted for the U.S. Department of Labor features both of these challenges in providing incentives: (i) addresses were not available for most sample members, which ruled out pre-incentives; and (ii) the instrument featured both a screener and an extended interview. The Employee Survey is a national dual frame RDD survey of adults employed in the last 12 months. Adults who needed or took family/medical leave in the 12 months prior to the interview

were oversampled and administered an interview roughly twice the length of the interview for respondents who did not need or take such leave. This presentation describes results from a randomized experiment to assess the impact of a post-paid incentive on cooperation rates, data quality and cost per completed interview. Special focus is given to the effect of the incentive on cooperation among cases receiving the longer questionnaire and cases in which the screener respondent was not the adult selected for the extended interview.

Using the iPad as a Prize-Based Incentive to Boost Response Rates: A Case Study at Brigham Young University

Richard McClendon, *Brigham Young University*; Danny Olsen, *Brigham Young University*

In 2009, Dillman, Smyth, and Christian downplayed the use of prize drawing incentives for Web-based surveys and instead conclude that, like mail and telephone surveys, the most effective way to increase response rates in Web-based surveys is to use postal mail to deliver an invitation and prepaid cash incentive (pgs. 274-275). However, for many public, marketing, and social researchers, the feasibility of this approach is not only cost-prohibitive but naturally goes against the initial purposes of using the Internet in the first place—the reduction in time and ease of use. Further, when it comes to the advancement and public use of technological, data from 2009 already feels like it's a century behind. Thus, the purpose of this paper is to revisit the question of lottery- or prize-based drawings, particularly in light of using new technological devices as incentives; in our case—the iPad3. During 2011 and 2012, the Office of Assessment and Analysis at Brigham Young University sent out several surveys to both students and alumni that included an iPad drawing as an incentive. Data gathered from these surveys clearly show a significant increase in response rates for both students and alumni. Some of these increases have ranged from 8% to 13%. Given these favorable increases compared to the relatively low cost of offering an iPad in a drawing, we feel this simple application would represent an attractive solution to maintaining a sustainable cost/benefit trajectory for future research and polling among other institutions. We will present further details surrounding this research including a discussion of demographics that identify who are more or less likely to respond to a survey that includes a drawing for an iPad.

Tracking Children Across Key Transitions Using Data from Multiple Informants—Lessons Learned from the Head Start Family and Child Experiences Survey

Annalee Kelly, *Mathematica Policy Research*; Marcia Comly Rigby, *Mathematica Policy Research*

Longitudinal studies of young children often focus on key transitions, such as the transition to school, with the goal of estimating characteristics of children before and after these transitions. The accuracy of these estimates depends in part on having high response rates across data collection waves, and expert tracking of study respondents is imperative. In addition, new sources of data are often needed as children transition from one program to another. The Head Start Family and Child Experiences Survey (FACES) is a national, longitudinal, descriptive study of children and families served by Head Start. It follows a national sample of children from Head Start entry through program participation and to the end of kindergarten. An accurate accounting of study children before, during, and after each round of data collection is necessary in order to guarantee the integrity of the sample. When children leave Head Start, their kindergarten teachers become an important source of information on their kindergarten programs and any difficulties they might be having at school. For the most recent FACES

cohort, FACES 2009, Mathematica tracked a sample of low-income, preschool-aged children from Head Start entry through the end of kindergarten. Schools and teachers were identified, located and confirmed, despite challenges that included contacting parents in hard-to-locate populations and identifying, introducing the study to, and gaining cooperation from kindergarten principals and teachers previously unconnected to the study. In a 16-week period, we identified 96 percent of children's kindergarten schools and 93 percent of their teachers. This methodological brief examines how we were successful in mitigating these challenges by using data from multiple informants (parents during interviews at the end of Head Start and again in kindergarten, Head Start programs, and elementary schools), and examines how useful each was in providing verifiable information to help locate children.

When is Enough Enough? Deciding the Optimal Number of Contacts for a Multi-Mode Survey

Kerry Levin, *Westat*; Jocelyn Newsome, *Westat*; Pat D. Brick, *Westat*; Brenda Schafer, *Internal Revenue Service*; Ron Hodge, *Internal Revenue Service*; Patrick Langetieg, *Internal Revenue Service*

A variety of factors can improve survey response rates, including incentives, a credible sponsor, and a brief, easy-to-complete survey. In addition, the number and form of contacts during survey administration can significantly influence response rates. The universally accepted procedures for conducting mixed mode surveys are based on variants of Dillman's Tailored Design Method (TDM) (Dillman, Smyth, and Christian 2009). The classic TDM approach advocates contacting respondents 4 or 5 times, where each successive contact is different from the preceding contact. It has been empirically demonstrated that each additional contact will result in an increase in the overall response rate (Hassol et al. 2003, Rookey et al. 2012). When plotted as a curve against level of effort or cost, response rates move incrementally towards an asymptote or a plateau. However, in actual survey practice, we rarely observe a plateauing of the response rate. For many reasons, including budgetary and time constraints, more than 5 contacts is typically not an option in "real world" survey practice. As a result, there is minimal evidence in the literature concerning the optimum number of times a respondent to a survey should be contacted in order to increase response rates and still be cost efficient. In other words, when is enough, enough? In this paper, we explore adding a sixth contact to the IRS Individual Taxpayer Burden (ITB) Survey, which includes a sample of taxpayers across the United States. The 2010 ITB Survey, which had five contacts, never reached a plateauing of response rate. Given that, a sixth contact was added to investigate whether this plateauing effect would be observed. We use three critical measures to determine the success of this sixth contact: the cost per additional complete, respondent feedback collected via a toll-free helpline, and response rate analysis at each of the phases of contact.

Incentives and Early-Life Civic Engagement as a Mediating Factor in a Study After 50 Years

Ashley Kaiser, *American Institutes for Research*; Danielle Battle, *American Institutes for Research*; Jizhi Zhang, *American Institutes for Research*

Civic engagement has been considered as one of the factors that influence individuals' survey participation (Singer et al. 1999). According to Groves et al. (2000), when sample members with higher levels of civic engagement were offered incentives for survey participation there was no impact on cooperation, but when sample members with low levels of civic engagement were

offered the same incentive, there was a positive effect on response. This paper examines the combined influences of individuals' civic engagement and cash incentives on survey participation. Utilizing the nationally representative data of Project Talent, this study explores the extent to which incentives affect sample members' response propensity, by controlling their early-life civic engagement. Project Talent (PT), a longitudinal study started in 1960, collected extensive cognitive, personality and background information from 440,000 9th-12th graders. Fifty years after the base-year data collection, a pilot test of one percent of the original sample was conducted in 2011. The pilot test involved an incentive experiment, with one-third of sample members not receiving any cash incentive, receiving a noncontingent \$2 bill, or receiving a noncontingent \$20 check. Civic engagement was measured through respondents' clubs/organization participation and political involvement in high school in 1960. This paper will use the items measuring civic engagement to examine their effect on response to a follow-up survey 50 years later. The findings of this paper will enable researchers to better understand the influence of early-life civic engagement and survey response propensity and determine whether high civic engagement in early-life incentivizes participants to participate in surveys later in life, regardless of monetary incentives.

Responsive Design Features and Respondent Cooperation in the Health and Retirement Study

Piotr Dworak, *University of Michigan*; Heidi Guyer, *University of Michigan*

Responsive design relies on observation of data collection progress and application of measures to increase cooperation thereby reducing non-response (Groves and Heeringa 2006). However, responsive design may have different implications for cross-sectional and longitudinal studies. Over the past waves, the Health and Retirement Study (HRS) operations team has implemented many targeted strategies to reduce the data collection timeline and secure cooperation of late respondents. The goal of this analysis is to assess the efficiency the responsive design features on instantaneous (within-wave) progress and the impact on longitudinal cooperation. Examples of such interventions include: reducing the length of the baseline interview, an experimental design randomizing wave 2 respondents into higher and lower incentive conditions (in some cases higher and lower than their wave 1 incentive), targeted "end-game" mailings, interviewer bonuses to prioritize different types of cases, "kept-appointment" incentives and emphasis on contacting respondents around the holidays. The goal of analysis is to comparatively assess these interventions and their impact on within-wave and longitudinal cooperation. The analysis will further the understanding of responsive design strategies and their application to the longitudinal studies.

Video Effects on Panelist Co-operation: Arbitron Installation Video

Kate T. Williams, *Arbitron*

Arbitron's Portable People Meter (PPM) is a device that automatically detects an individual's exposure to encoded media and transmits the data to Arbitron for reporting. Households are recruited into a two-year panel, and their members are asked to wear the PPM from the time they rise in the morning until they retire at night. In order to comply successfully, the household must first install the PPM equipment. Due to rising costs associated with recruiting households, Arbitron is exploring methods to improve panelist installation rate. Across 2012, Arbitron conducted a series of trials with newly recruited PPM households to examine the effect of an email containing an offer to view an installation video. The installation video provided guidance

on how to set up PPM equipment, and it was available on a website that panelists could access only through the email link. In this experiment, approximately half of the newly recruited households received the email with a link to the installation video; the other half received a similar email without such a link. Differences in the graphical content and the subject lines of the emails were also tested. Analyses of households' behavior after receiving the email focus on installation success rate, and also include the efficacy of different email communications.

Innovative Measurement of Public Opinion

140 Characters or Less to Shape Public Opinion: Methodological and Theoretical Improvements on the Use of Twitter to Measure Public Attitudes **Anna Novikova, *Knox College***

How can social media complement traditional surveys in assessing public opinion? While telephone surveys constrict responses (and introduce bias), moving from asking to listening allows us an unfiltered look at public opinion. Translating open-ended opinions into useful figures, however, is a challenge. Using a corpus of English language tweets from a 1% sample stream of public Twitter posts collected in the two months prior to the 2012 presidential election, I assess the validity of using Twitter as a forecasting tool. A machine learning algorithm trained on hand-coded data is used to measure sentiment (i.e. positive and negative emotions) expressed in the Twitter data. I aggregate these sentiment scores and compare them to public polling data within the same time frame. I build upon previous research in this area by using more sophisticated classification techniques, rather than either naïve volume counts or list-based classification. In this way, what is being said about a candidate is captured, rather than how often a candidate is mentioned. I hypothesize that opinions expressed by Twitter users, who are more educated and more informed than the general public, will be more responsive to day-to-day events in the course of the campaign. Changes in sentiment among these users, then, should be a leading indicator for movement in public polls. This research contributes to the development of social media analysis as a supplement to traditional public opinion polling.

Understanding Elections: Voter Intentions, Expectations, and Forecasts **David Rothschild, *Microsoft Research***

Using a unique dataset from YouGov/Xbox Polls we explore the relationship between respondents' intentions and expectations. During the 2012 election Xbox conducted roughly 750,000 interviews with 350,000 respondents. These respondents answered questions about their candidate support and engagement in the election, as well as their expectation of who would win, who their social network supported, and who the media was projecting to win. Cross-sectional analysis of how intentions relates to expectations explains the underlying structure of how respondents view the election. Panel analysis of how intentions and expectations move during the election cycle provides new insight into the bandwagon effect of expectations effect on intentions. Finally, we show how aggregations of respondents' expectations accurately predict both national and state-by-state elections.

Wanted: Young Adults 18-35 – Leveraging Smartphone Applications for Repeated Measures of This Elusive Cohort

Shu Duan, *The Nielsen Company*

Growing smartphone penetration has offered survey researchers a new mode to reach the young adults for measurement. Latest research (Pew Internet Project, Sep 2012) shows that 2 out of 3 under aged 30 adults own a smartphone which reveals a strong potential in using smartphone to reach the younger cohort that is usually hard to reach by traditional survey methods. Past studies have focused on specific areas of mobile research such as cell phone frame, survey design on mobile browser, survey administration via text messages, etc. The research gap remains on the effective mobile research methodology to target young adults age 18 to 35. Nielsen will be conducting a pilot on crowdsourcing from a mobile panel to collect media consumption through a smartphone application with the emphasis in researching respondents under age 35. Specifically, we would study 1) respondent cooperation through crowdsourcing from a mobile panel; 2) app usability optimized for survey data collection; and 3) survey compliance of reporting media consumption in smartphone app. This study will share the learning on the end-to-end methodology of leveraging a mobile panel of smartphone users to gain cooperation from young adults age 18-35 and adapting the smartphone features for respondent engagement to maximize their participation throughout the data collection period.

Enhancing Usability and Data Quality

Usability of App Features and Tutorials

Kelly L. Bristol, *The Nielsen Company*; Jennie Lai, *The Nielsen Company*; Michael W. Link, *The Nielsen Company*

A critical question about the sustainable future of survey research is how to design an effective user experience for electronic data collection tools. Usability research defines a well-designed user experience as being easy to use, quick to complete, memorable, with minimal errors and well-liked by users. Optimizing the user experience reduces respondent burden and can significantly improve data quality. Developing a user centered design is particularly important for long term panel and diary studies where respondents must interact with the data collection instrument frequently for an extended period of time. Findings reported here assess the usability of features in a mobile and Web app for a two week diary study of television viewing conducted by Nielsen in August of 2012. Within the application there are four primary modules – enter viewing, check entries, messages and badges. In addition, there is a tutorial feature for each module and the home page. Usability of the application modules is measured on several different metrics depending on the module purpose and depth of features. Ease of use and quickness are evaluated by time-to-complete surveys on the first use compared to the average completion time for the overall study. The effectiveness of the tutorial feature is also evaluated through two forms of comparison: 1) pre and post tutorial usage of features, 2) tutorial versus non-tutorial user survey completion times and feature usage. Likability measures are supplemented from a post-study survey. This research provides insight into developing effective user experience design for a self-reporting electronic data collection tool, and the effectiveness of app tutorials on optimizing user experience.

From 1.0 to 2.0: Lessons Learned of Mobile Application Design for Effective Respondent Engagement

Jennie W. Lai, *The Nielsen Company*; Kelly Bristol, *The Nielsen Company*; Michael W. Link, *The Nielsen Company*; Shu Duan, *The Nielsen Company*

The continued surge of smartphone ownership and mobile application (app) usage has opened doors for survey researchers to reach the young adults and ethnic minorities. Using mobile apps as the data collection tool and the versatility of its features allow for new respondent engagement techniques unparalleled to the traditional modes of data collection. Both user interface and user experience design of the mobile app are the core tools for user engagement and the key to encourage compliance throughout the data collection period. Mobile app features such as dynamic tutorial for survey instructions, in-app notification for customized respondent communication, deployment of badges as incentive for survey compliance, social sharing through Facebook posting, etc. are the tools designed to keep respondents engaged for repeated measures. Nielsen has conducted two pilots in January and August of 2012 to capture media usage behavior through two comparable versions of mobile application. The latest mobile app study was launched in two markets using dual telephone frame sample for recruitment and respondents participated for a two-week collection period. The first pilot yielded insightful learning on the effectiveness of the aforementioned mobile app features for respondent engagement and significant app enhancements were made for the second pilot. This research paper will discuss the lessons learned of the app features from the first pilot and compare the results of the upgraded features in the second pilot. The findings of these research studies will inform which mobile app features hold promise for respondent engagement targeted for repeated measures of longer term panel studies.

Can Embedded Help Text Links in Web Survey Items Improve Data Quality?

Natasha Janson, *RTI International*; Christopher Bennett, *RTI International*; Lesa Caves, *RTI International*; Melissa Cominole, *RTI International*; Bryan Shepherd, *RTI International*; Jennifer Wine, *RTI International*

Self-administered surveys often include text that is separate from survey items and serves to provide respondents with standardized definitions and clarifications for nuanced items. For Web-based surveys, this information can be presented in a variety of formats, including “Help” buttons leading to external Websites or popup windows. More information is needed to evaluate the extent to which these various formats for accessing help text actually encourage its use and whether the use of help text has any effect on the responses provided. Embedded help text links were evaluated in two large postsecondary surveys. Help text in these surveys has historically been accessible via a “Help” button on each survey form, and has generally exhibited very low usage rates among self-administered respondents (typically about one percent). To make the help text feature more salient for self-administered respondents, key words were hyperlinked so that respondents could click on the linked words and access the help text for that form, just as if they had clicked the “Help” button. The content was the same regardless of how help text was accessed. The embedded help text links were used only on selected survey items, while all forms displayed the “Help” button at the bottom of the form. Preliminary results show the use of help text increased significantly on screens with embedded links versus screens with only the separate “Help” button. Implications for survey timing and response distributions will be discussed. Study findings indicate that the way in which help text is presented has implications for Web survey administration and data quality.

Grid Formats, Data Quality, and Mobile Device Use: A Questionnaire Design Approach

Colleen A. McClain, *Survey Sciences Group, LLC*
Scott D. Crawford, *Survey Sciences Group, LLC*

Grids have been the subject of significant research as a frequently used—but often problematic—way to present multiple questions in a shared layout, particularly within Web-based surveys. Respondents' increasing use of mobile devices underscores and emphasizes the need to reexamine design standards for grids and questionnaires that will now be seen on a variety of screen types. While recent work has begun to explore the relationship between device use, data quality (McClain, Crawford, & Dugan, 2012; Saunders et al., 2012), and substantive responses (Mavletova & Couper, 2012), considerable practical concerns remain in conducting surveys that have been optimized for larger screens. Drawing upon recent literature and paradata that we have collected, we propose a combined layout and questionnaire design approach to confronting these challenges-- acknowledging that while refining layout and user design of grids can impact data quality (Couper et al., 2013) and aid mobile navigation, an additional challenge lies in designing questionnaires that are clear, cohesive, and adaptable to the smaller screen space available on mobile devices. To better understand interactions between device use and data quality measures in a grid-heavy setting, we reviewed respondent behavior and characteristics of grids from multi-year administrations of 11 Web surveys with college student populations, spanning several hundred thousand respondents. We focused our exploration on key contextual characteristics of grids that may influence data quality and exacerbate burden--such as questionnaire position/context, grid length and density, scale design, sensitivity of content, and presence of validations. Specifically, we investigate the relationship between several of these characteristics and mobile respondents' tendency to straightline, as a potential indicator of satisficing (Krosnick, 1991); to break off; and to yield higher rates of item-missing data. Our presentation will highlight key findings from this analysis and discuss implications for questionnaire design that considers the mobile space.

Examination of Question Complexity Through Paradata

Rebecca J. Powell, *University of Nebraska-Lincoln*; Ana Lucia Cordova Cazar, *University of Nebraska-Lincoln*; Jinyoung Lee, *University of Nebraska-Lincoln*

Question complexity in surveys should be at a level where all respondents can understand what the question is asking (Dillman et al. 2009; Groves et al. 2009). Therefore, in practice, researchers aim to create questions that are no higher than an eighth grade reading level. While this gives a quantitative measure for the overall question, there can still be qualitative aspects of a question that make it complex even when the reading level is below eighth grade. For example, a question can be phrased such that it is below an eighth grade reading level, but the ambiguity of the words in the question can lead to a complex question. Programs like QUAID help to point out these challenging words and phrases, which can lead to difficulty with the response process. When respondents have difficulty with any phase of the response process, it can have adverse effects on data quality. One way to test the effects on data quality is through paradata. Specifically, paradata allows us to collect the frequency of answer changes to questions, and back-ups where respondents answer another question before changing their answer to a previous question. This study uses the Internet component of the Gallup Panel to develop a question complexity index from QUAID information, the question reading level and word count. These are then examined to better understand the relationship between question

complexity and the frequency of answer changes and back-ups per question. Preliminary findings show a 0.36 correlation between the reading level and the average number of answer changes but a 0.53 correlation between the word count and the average number of answer changes. Increased answer changes can result in measurement error if respondents are unsure of their answers to questions.

Using Mail to Improve the Effectiveness of Web and Telephone Data Collection for Address-Based Samples of the General Public

Using Visual Design to Aid Within-Household Selection in Mail Surveys: Does it Lead to Accurate Selection and Representative Samples?

Mathew S. Stange, *University of Nebraska*

Research examining the next and last birthday methods of within-household selection in mail surveys find few differences in sample composition between the two methods, but find both methods are unrepresentative of certain demographic groups (e.g., Battaglia et al. 2008). Yet other research shows that accurate selection of respondents remains a problem for within-household selection in mail surveys (e.g., Olson & Smyth forthcoming), with inaccuracy rates ranging from a small percent to over 30% (e.g., Battaglia et al. 2008; Schnell 2007). Because interviewers are not present, mail surveys require a different approach to motivate within-household selection and to aid households in selecting the correct household member. Visual design is one way to possibly help. In this study, we examine the use of a calendar placed on a survey's cover letter to help households select the correct household member with the next birthday. Including a calendar adds emphasis to the task and may aid households in selecting the correct respondent. Data come from the 2012 Nebraska Annual Social Indicators Survey (NASIS; n=959, AAPOR RR1 26.6%) – an omnibus mail survey of Nebraskans. Half of sampled households received a cover letter with the calendar and the other half received a cover letter without the calendar. We examine the resulting sample composition and use a household roster included in all the surveys to evaluate the accuracy of selecting the household member with the next birthday. Preliminary analyses indicate that the response rate did not differ significantly between the treatments (26.5% with calendar; 26.7% without calendar) and the sample had similar representation on education levels. We also examine whether the calendar increased accuracy of within-household selections, using the 92% of the sample who completed at least some information in the roster. We conclude with implications for within-household selection methods in mail surveys.

Effects of Survey Sponsorship on Internet and Mail Response: Using Address-Based Sampling

Michelle L. Edwards, *Washington State University*

Scholars have shown that the combined use of token cash incentives with an initial withholding of a mail response alternative can increase Internet response rates significantly in regional and state-level surveys using address-based sampling. However, the effectiveness of this model has declined when university sponsors have surveyed residents in distant states. While nonresponse rates are not necessarily predictive of nonresponse bias, attitudes toward a

survey's sponsoring organization may influence both response rates and nonresponse bias. To test the effects of survey sponsorship by a local (in-state) university sponsor versus a distant (out-of-state) university sponsor on response rates, we conducted an experiment in spring 2012 with an address-based sample of Washington and Nebraska residents. We found that sponsorship had a significant effect on final response rates in both states, with in-state sponsorship significantly improving response for both mail-only and 2 Web+mail (initial Web request with a mail questionnaire offered in the fourth and final contact) treatment groups. For 2 Web+mail groups, we also found that local sponsorship increased the risk of responding by Web (relative to not responding), but not the risk of responding by mail (relative to not responding). In examining the representativeness of the resulting samples, we found that our survey respondents were both generally older and more highly educated than state-level estimates from the Gallup Poll and American Community Survey. In Nebraska, a Republican-leaning state, distant-sponsored surveys obtained a lower percentage of Republicans than local-sponsored surveys. In Washington, a Democrat-leaning state, local-sponsored surveys obtained a lower percentage of Republicans than distant-sponsored surveys. This research suggests that recent public opinion findings demonstrating declining public trust in science among conservatives (but not other groups) may have important consequences for university-sponsored survey research.

Sample Performance and Cost in a Two-stage ABS Design with Telephone Interviewing

W. Sherman Edwards, *Westat*

Random-digit-dial (RDD) surveys long provided an effective, lower-cost alternative to face-to-face surveys for general population research. With declining response rates and an increasing proportion of cell-phone-only households, both the effectiveness and cost of RDD surveys have become less attractive. Address-based sampling (ABS) is becoming a preferred approach in many cases, but there is no consensus as yet on the optimal data collection mode or mix of modes, particularly for surveys requiring within-household sampling and/or an interviewer-mediated questionnaire. Brick et al. (2011) describe a successful two-stage mail ABS design, where the first stage determines household eligibility and provides information needed for within-household sampling, and the second stage collects more detailed information about sampled individuals. Two-stage designs have also incorporated telephone interviewing at the second stage. This paper will present the results of a pilot two-stage ABS design for a companion survey to the National Crime Victimization Survey to support local area estimates, with an initial mail contact and telephone interviewing. The pilot incorporates a split-sample experiment. In one treatment, only addresses without an associated telephone number were sent the mail instrument, with the objective of obtaining a telephone number. In the other treatment, all sampled addresses were sent the mail instrument, which also included questions to allow stratification of the sample by likelihood of having experienced a crime. In both treatments, telephone interviews were attempted with all households for which a telephone number was obtained. The analysis will compare sample performance and per-case cost between the two treatments and with the likely sample performance and cost of an RDD survey to accomplish the same objectives. Since the NCVS estimates both prevalence and characteristics of relatively rare events (crimes), a large sample is required. Therefore, we will calculate both costs per completed interview cost per completed interview with reported victimization.

Is Pushing the General Public to the Web in Address-Based Samples Cost Effective?

Virginia M. Lesser, *Oregon State University Department of Statistics*

Interest in using mail contact in address-based samples of the general public to encourage responses over the Internet is considerable. However, several studies have shown that it is necessary to also use mail questionnaires in order to obtain responses from households with quite different demographics than those who will respond by Web (e.g. Messer and Dillman, *Public Opinion Quarterly*, 2011). That study also shows that “pushing” some respondents to the Web may actually increase total survey costs on a per respondent basis while reducing overall response rates and not provide a demonstrable improvement in household representation. The expected savings from questionnaire mailing and processing costs did not offset the set-up and implementation costs. In this paper we examine results from two experiments conducted on address-based samples in Oregon during 2010 and 2012. Response rates and cost effects of two approaches were examined: 1) Web+mail (withholding mail in early contacts) and 2) offering a choice of Web or mail were compared using a mail-only approach as controls. We systematically examine for each approach and year response rates, costs for each survey mode, and demographic representation with regard to age, gender, and employment. Thus, we reexamine the question of whether including a Web response is cost effective when administered in a somewhat different way that used by Messer and Dillman.

Using GIS to Target Address-Based Samples of Households for a Web (vs. Mail) Response: Evidence from Three Web+Mail Surveys in Washington State **Benjamin L. Messer, *Washington State University***

Address-based sampling enables researchers to use geographic information systems (GIS) to analyze the social, demographic, and other characteristics of the communities in which sampled households are located. Increasingly, research is finding that these methods are important for survey designs in which households can be targeted for response to different survey modes in advance of the data collection period. However, little is known about which community characteristics are important for predicting what households have the highest propensities for responding to a Web (vs. mail) survey. Previous research has identified a number of individual and household characteristics that are important for predicting Web response, including household Internet access, socioeconomic status, and age, but less attention has been directed toward the community-level. The purpose of this paper is to report on the geographic bases of Web and mail survey response to statewide surveys, identify those characteristics that are most salient for targeting households to respond via the Web, and to offer suggestions on which Web+mail methods may be the most effective in different types of communities. We use existing data of address-based samples from three general public Web+mail surveys conducted in Washington State between 2008 -2011 matched with data from the Census and American Community Surveys in GIS. Analyses are currently being conducted and results will be available in the next few weeks.

Public Opinion and the Environment

The Weathering of Skepticism: An Examination of American Views on the Existence of Climate Change

Christopher P. Borick, *Muhlenberg College Institute of Public Opinion*; Barry G. Rabe, *University of Michigan*

The period between 2008 and 2012 was one of significant shifts in American public opinion regarding climate change. Between 2008 and 2010 an increasing number of Americans indicated skepticism that global warming was occurring. This trend has been reversed between 2010 and 2012, with most public opinion research finding that levels of acceptance regarding the existence of global warming returning to levels observed in 2008. Numerous studies have identified factors such as changing economic conditions, media framing and variations in weather as the determinants of the shifts in American public perceptions about climate change. In this study we examine the role that individual perceptions about weather have had on their beliefs regarding the planet's climate. In particular we look at the personal experiences that Americans have had with conditions such as severe droughts, hurricanes and heat waves, and how those experiences have diminished skepticism regarding global warming. The study includes results from 9 iterations of the National Survey of American Public Opinion on Climate Change (NSAPOCC) between 2008 and 2012, including a rounds conducted just before and after Hurricane Sandy's landfall in late October of 2012.

Global Warming Attitudes Among Local News Viewers and Non-Viewers; Media Market Comparative Analysis and Change Over Time

Amy Simon, *Goodwin Simon Strategic Research*; Leora Lawton, *Tech Society Research*; UC Berkeley, *Berkeley Population Center*; Adam D. Probolosky, *Probolosky Research LLC*; Paul A. Hanle, *Climate Central*

In light of AAPOR's conference theme "Toward a Sustainable Future for Public Opinion and Social Research" we submit a paper looking at views and attitudes about global warming and climate change. This paper reports on the findings of two surveys measuring attitudes and views towards global warming in three media markets, comparatively, as well as over time. In February 2012, we completed a benchmark telephone survey with n=6,089 completed interviews using live interviewers in three media markets (DMAs): Denver, Colorado; Terre Haute, Indiana; and Dallas, Texas. We conducted approximately 1,000 interviews in each market among adults who watched a local news station at selected evening viewing times that included the weather report, with a focus on a different network affiliate in each market. For a control group, we also conducted 1,000 interviews in each market among adults who either did not watch the targeted local news station or watched the station at different times than the select evening viewing times. The RDD sample included landlines and cell phones. In our benchmark survey, we found that while six in ten respondents think global warming is happening, just over four in ten are concerned about its impact on the world today. By combining an attitudinal survey with media consumption, we were able to show that the source of information about global warming as well as religious and political ideological positions are strongly associated with attitudes about global warming, and that these positions are independent of educational attainment. In February 2013, one year later, we will conduct the survey in the same markets to measure any change in attitudes over time among the viewer

and non-viewer populations. We will also investigate whether watching certain weather newscasts has a quantifiable impact on views of global warming.

Polls, Publics and Pipelines: Mapping Public Opinion Toward the Keystone XL Pipeline in the United States and the Northern Gateway Pipeline in Canada
Timothy B. Gravelle, *PriceMetrix Inc.*

The politics of oil pipelines has been especially prominent in recent years in North America. In the American case, debates about economic benefits, energy security and environmental impact have been provoked by the then-proposed (and now vetoed) Keystone XL pipeline intended to take bitumen from northern Alberta in Canada to refineries on the Gulf of Mexico in Texas. In the Canadian case, similar debates have been provoked by the proposed Northern Gateway Pipeline from northern Alberta westward to ports in British Columbia. Drawing on data from recent probability-based surveys in the U.S. (by the Pew Research Center) and Canada (by Ekos Research Associates), this paper asks a series of questions comparing the two cases. What levels of support for (and opposition to) the two pipelines exist? What are the roles of political factors (such as party identification), economic attitudes and proximity to the proposed pipeline routes in shaping attitudes? And how do political and economic factors (on the one hand) and proximity to the pipelines (on the other) interact? In asking these questions, the paper sets out to build on the growing body of literature highlighting the geospatial determinants of policy attitudes.

Emphasis Framing and Americans' Perception of Scientific Consensus: Scientists Agree on "Climate Change" but not on "Global Warming"
Jonathon P. Schuldt, *Cornell University*; Sungjong Roh, *Cornell University*; Norbert Schwarz, *University of Michigan*

Whether or not citizens perceive a scientific consensus on global climate change has emerged as an important factor in public opinion regarding climate policy (Weingart, Engels, & Pansegrau, 2000; Kahan, Jenkins-Smith, & Braman, 2011). However, little is known about the situational factors that might influence this perception. Building on recent research (Schuldt, Konrath, & Schwarz, 2011), we explore whether a seemingly trivial wording change can influence perceptions of scientific consensus, namely, whether the issue is framed in terms of "global warming" or "climate change" in the survey question. In a nationally representative survey experiment (N = 2041) fielded August 25–September 5, 2012, respondents reported on their own as well as scientists' beliefs about the existence of global climate change, worded either in terms of global warming or climate change. Replicating a previous observation (Schuldt et al., 2011) with a representative sample, Republicans (but not Democrats) reported significantly lower existence beliefs when asked about "global warming" as compared to "climate change." Going beyond their own beliefs, respondents overall were less likely to perceive scientific consensus when the issue was framed in terms of global warming. Thus, the influence of these emphasis frames, which are commonly used interchangeably in public discourse, extends beyond personal beliefs and affects citizens' perceptions of the positions of scientific experts. Discussion focuses on theoretical and practical implications of this subtle but overlooked factor in science communication, survey design, and public opinion about climate.

Global Warming, Geo-Engineering and Human Happiness: Survey Based Estimates of Worldwide Gains and Losses in North and South, Winter and Summer

Jonathan Kelley, *International Survey Center and University of Nevada, Reno*

This paper provides quantitative estimates of the consequences of global warming for human happiness (well-being, utility, life satisfaction). Data are from a representative national sample of the U.S. (N=2295), together with standard NOAA data on climate worldwide on a half-degree latitude/longitude grid. Regression estimates show that a century of warming at currently expected rates will increase American's satisfaction with winter weather in northern and mid-latitude states but decrease their satisfaction with summer weather in all states. The gain is equivalent to that which would come from an increase in income of around 8% in northern states and a loss of 5% in southern states—huge figures, dwarfing most other consequences of climate change. Assuming people in other nations evaluate temperatures the same way as Americans, global warming is likely to be beneficial in higher latitudes (Canada, northern Europe, north China, Korea, Argentina, New Zealand) and bad near the equator (Mexico, Central America, Brazil, sub-Saharan Africa, India, south China, south-east Asia). The potential for North-South conflict is clear. Moreover, choice in these matters may not lie with western nations: In the absence of geo-engineering, the continued expansion of coal fired power plants is likely to be a benefit to the north Chinese and possibly to China as a whole. If so, and if the large and rapidly growing Chinese economy pursues its own self-interest, that alone could lead to global warming regardless of what policies western nations pursue at home. Geo-engineering techniques (such as atmospheric sulfur injections) might perhaps reduce these conflicts by cooling lands near the equator while letting temperatures rise at higher latitudes. Indeed geo-engineering might be a worldwide benefit if it could selectively cool summer temperatures in middle and lower latitudes while letting winter temperatures rise at middle and higher latitudes.

Panel Recruitment, Attrition and Data Quality I

Predicting Survey Breakoff in Internet Survey Panels

Tarek Al Baghal, *University of Nebraska - Lincoln*; Allan L. McCutcheon, *University of Nebraska - Lincoln*; Davit Tsabutashvili, *University of Nebraska - Lincoln*

Survey breakoff – when respondents discontinue their participation before completing the questionnaire – has attracted a growing amount of interest and attention (see, e.g., Peytchev 2009). The increased interest in breakoff involving Internet survey respondents has been accelerated by the relatively recent availability of paradata collection methods for Web surveys. 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 to questions, time of day when the survey breakoff occurs, as well as a number of other factors that can be evaluated as contributors to survey breakoff. The proposed study examines data from monthly waves of the Internet component of the Gallup Panel, a multi-mode (mail and Web) panel of American households. In addition to standard demographic respondent characteristics and survey design factors (e.g., question complexity, topic, number of questions on the page, length of survey), the analysis will include a variety of respondent self-reports on Internet sophistication and survey design factors and paradata to explore factors related to survey breakoff. Preliminary analysis indicates that while long-term panel members are less

likely to breakoff, that 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 responsive/adaptive design (Groves and Heeringa 2006) interventions for Internet surveys that may prove useful in averting, or delaying, Internet survey breakoff.

Innovative Retention Methods in Panel Research: Can SmartPhones Improve Long-Term Panel Participation?

James J. Dayton, *ICF*; Andrew Dyer, *ICF*

Minimizing participant attrition is vital to the success of longitudinal panel research. One such example of 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. Specifically, the NRBS Program enables the Coast Guard to better identify safety priorities and coordinate and focus research efforts. The project features a several components, one such being the "Trip Panel." The Trip Panel is designed to capture actual exposure to recreational boating. This panel was recruited via dual-frame, dual-mode (Random Digit Dial telephone and mail) and has been in place for over a year. Respondent contact information includes e-mail address, mailing address, and telephone. In many cases, the provided contact number is a mobile device. This presentation will explore ICF researchers' quest to improve panel retention through the introduction of a smartphone application that engages respondents in between survey waves by allowing them to communicate changes in contact information and even provide survey responses via smartphone rather than via the Web or traditional telephone. Active panelists who provide cell phone contact information will be randomly assigned to receive standard retention communications via mail, phone and e-mail (control) or alternate retention communications via a smartphone application and text message/SMS (treatment). The communications application for the treatment group includes study updates, various interactive communications, and mini-surveys. ICF researchers will analyze the differences in control and treatment panel retention over a six-month period. We will also survey panelists' willingness to sign-on for another annual wave of the panel as well as their overall satisfaction with panel participation as an indicator of long-term continued participation.

Probability Based Postal Recruitment into Longitudinal Online Panels: The Effects of Personalization and Incentives

Johan Martinsson, *University of Gothenburg*

This study examines the feasibility of probability based recruitment into longitudinal on-line panels through postal invitations. The study explores the effect of three factors: personalization, incentives and reminders. Further, the study uses a factorial design allowing us to explore interactions between for example incentives and personalization. The aim of the study is to find the most cost-efficient way to recruit a reasonably representative probability sample. Since this large scale study involves as many as 29 000 post cards being sent to a probability sample of the Swedish population from the national population register, we are able to analyze both the effect on the recruitment rate of personalization, incentives and of reminders, but also the effect on the actual demographic and attitudinal representativeness of those recruited from different kinds of postal invitations. Further, due to the excellent Swedish population register we also

have access to register data on marital status, age, sex, children, citizenship, country of origin and more for all individuals included in our random samples, and not only in the aggregate. This allows us to carefully check which demographic groups respond stronger (or weaker) to the factors examined in this study. All in all, three main outcomes are examined: recruitment rates, representativeness, and the cost of recruitment. Finally, we also check the long-term effect of different recruitments after the respondents participate in their first large scale survey after the initial recruitment survey approximately one month later.

Acquiescence to False Preload Information When Using Dependent Interviewing **Johannes Eggs, *Institute for Employment Research*; Annette Jäckle, *Institute for Social and Economic Research***

With Proactive Dependent Interviewing (PDI), respondents are reminded of the answer to a survey question they gave in a previous interview. The previous information is used to verify whether the respondent's status has changed, or as a starting point for asking about events since the previous interview. In either case, concern is frequently voiced that measurement error from the previous wave will be carried forward into future waves of the survey. In this paper we use data from the panel survey "Labour Market and Social Security" (PASS), linked to individual administrative records, to examine possible causes acquiescence to false preload information. During the interviews for wave 4 of PASS, the preload was faultily generated for a subgroup of 393 respondents regarding welfare receipt and respondents were given questions with false preload information. Only a part of the respondents contradicted the false preload. However, the error allows us to exploit a rare research opportunity to address the following questions: 1) To what extent do respondents confirm previous information when that is false? 2) How much of the apparent false confirmation is in fact due to false reporting at the previous wave of the survey? 3) To what extent is the false confirmation carried forward into the next wave of the survey? 4) To what extent can the acquiescence be explained by personal traits, response strategies, response difficulty, or interviewer characteristics?

How am I Doing? The Effects of Gamification and Social Sharing on User Engagement

Oana M. Dan, *The Nielsen Company*; Jennie W. Lai, *The Nielsen Company*

Gaming mechanics and concepts ("gamification"), as well as virtual "sharing" within social networks, are emerging tools to increase participation in surveys and especially to maintain cooperation in longitudinal studies. As customizable and personalized devices germane to respondents' environment and lifestyle, mobile devices have greatly facilitated the development of interactive measurement instruments that are able to challenge respondents, to evaluate and reward their behavior, and to broadcast it to others in real time. However, the mechanisms underlying the effects of gamification and social sharing on respondent engagement have not been fully unpacked. These mechanisms may be active (extroverted interaction or competition with other participants) or reflexive (introverted evaluation of one's own performance). This paper assesses these two mechanisms, relying on data from a 6-week study of an innovative mobile application to measure media consumption behavior. The iPhone application allowed users to record what they watched on TV, to earn badges and "ranks" based on their engagement with the app's various features, and to share their accomplishments with other users. Mixed-effects panel models show that self-evaluation (checking how one is doing) and positive reinforcement from others increase engagement, whereas extroverted competitive

interactions (sharing one's performance with other users) decrease it. These results are significant among the two groups of study participants: one that was gradually exposed to the gamification and social sharing features; and the other exposed to the full-featured app from the beginning. Gamification and social sharing have stronger positive effects for those who were gradually exposed to these features, showing that these effects are independent of other factors, and that they could be explained in part by the novelty of these features. This suggests that gamification and social sharing are effective and self-sustaining (hence, cost-efficient) incentives in panel studies, especially if they promote self-evaluation and keep the study exciting.

Evaluating Address-Based Samples I

The Implications of Excluding Inactive Mailing Addresses From ABS Frames

Rachel Harter, *RTI International*; Bonnie Shook-Sa, *RTI International*; Joseph McMichael, *RTI International*; Jamie Ridenhour, *RTI International*

Unoccupied addresses in address-based sampling (ABS) frames lead to inefficiencies in data collection and increased data collection costs. Some studies remove addresses flagged as vacant or new construction to improve efficiency and reduce data collection costs. However housing units that are vacant or under construction in the frame have the potential to become occupied and part of the eligible population for the survey. The longer the time lag between frame construction and data collection, the greater the risk that the flags are outdated. Thus there are tradeoffs between ABS sample frame coverage of the U.S. housing unit population and the efficiency of data collection, with the element of time shifting the balance. This paper explores the tradeoffs in the context of the U.S.P.S. Computerized Delivery Sequence file (CDS), which is often used as an address frame for surveys and whose coverage of the housing unit population has been researched. Sometimes the CDS is supplemented with traditional field enumeration or ABS frame supplementation methods such as CHUM to improve coverage, especially in areas that do not have city-style addresses. Recently the No-Stat file (NS) containing drop units, throwbacks, and addresses on contract carrier routes not receiving mail has been made publicly available, and it, too, has been used to supplement the CDS. This paper examines vacancy and new construction status in the CDS/NS files, the typical durations for housing units being flagged as vacant or new, the clustering of flagged addresses within geographies and within buildings, and the extent to which addresses move from the NS to the CDS file, or vice versa. With this information, survey designers can make a more informed decision whether to supplement the active housing units in the CDS/NS files with those flagged as vacant or under construction.

The Trajectory of the USPS DSF: Change in National Coverage for In-Person Interviewing 2000-2010

Colm O'Muircheartaigh, *NORC at the University of Chicago*; Ned English, *NORC at the University of Chicago*

Our continuing research program at NORC indicates that the proportion of the USA that requires in-field listing has changed substantially over the past decade, shrinking from 28% to 15% of the population; the United States Postal Service (Computerized) Delivery Sequence File ((C)DSF) provides a preferable alternative from a cost and efficiency perspective for the rest of the population. We use data from the NORC National Master Sample in both the 2000 and 2010

decades, which has listings for national surveys across environments and geographies in the USA, so show the depth and breadth of changes to the DSF over the past decade. Improvements in the CDSF have not been evenly distributed across the population, however, with some areas remaining static since 2002 and others that formerly required in-field listing now suitable for using the CDSF. Our paper examines the kinds of places that experienced the most change in CDSF coverage during the period in which the list underwent the most research with respect to surveys, e.g. the 2002-2012 decade. We will describe which micropolitan statistical areas have improved faster than average, and what the structural implications of such changes might be. Multimode Address-Based Sampling (ABS) also requires standardized addresses, which are often not available for sparsely populated areas and for undifferentiated apartment addresses within buildings. By examining the trajectory of change, we predict the future requirements for in-person surveys and for multimode ABS.

Building a More Powerful Model to Predict Areas Where USPSBased Address Lists May Be Used in Place of Traditional Listing

Frost A. Hubbard, Survey Research Center, *University of Michigan*; James R. Wagner, Survey Research Center, *University of Michigan*; Haoyu Gu, Survey Research Center, *University of Michigan*; Wen Chang, Survey Research Center, *University of Michigan*

Traditional field listing is an expensive method for obtaining high levels of coverage on area probability studies. Over the past decade, many studies have shown how using the U.S. Postal Service Delivery Sequence File (DSF) as a sampling frame for area segments, typically clusters of Census blocks, can greatly reduce costs while maintaining relatively high levels of coverage. In general, rural areas have lower levels of coverage than suburban or urban areas. However, this generalization is not uniformly true. Brick and colleagues (2011) devised a model that improved the prediction of areas which are likely to be well-covered by the DSF that includes many other predictors. Their prediction model was built using mainly American Community Survey data, on a relatively small scale and not using a nationally representative sample of area segments. Since new data are available from the 2010 Census, and since the National Survey of Family Growth (NSFG) uses a nationally representative sample of area segments in which the DSF listings are reviewed for correctness, we have the basis to develop an improved model. We will use Census 2010 variables, variables from the Census hard to count data file, and data on the DSF as predictors. Results from an experiment using this model in production will be presented.

Growing Survey Response Rates on Trees: Evaluation of Response Propensity Models Based on Logistic Regression Models and Random Forests Using Block-Group Information Appended to an ABS Sampling Frame

Trent D. Buskirk, *The Nielsen Company*; Anh Thu Burks, *The Nielsen Company*; Brady T. West, *Institute for Social Research, University of Michigan*

Address based sampling (ABS) enables survey researchers and statisticians to append a vast array of ancillary information to the sampling frame at the block-group level for virtually every sampling unit. Information such as median household income, percentage of renters, or percentage of householders over 55 can be used a priori as part of the sampling design or post-sampling to either improve the survey recruitment processes or serve as the basis for nonresponse adjustments. In this presentation we report the results of a study aimed at evaluating the use of a series of variables available both at the block-group and ZIP-code+4

levels from both the 2000 Census and other commercial sources to estimate response propensities for a national media diary survey (MDS). The MDS sample consisted of over 650,000 addresses randomly selected from a national ABS sampling frame. The response propensity models were constructed from a catalogue of over 50 ancillary variables using both random forests and logistic regression models incorporating principal components for reduction of the ancillary data. These methods will be compared to a basic response propensity model derived using logistic regression from household predictors including age and Hispanic indicators. We first compare the internal validity of these models, derived using a series of cross-validation techniques including bootstrap resampling and a test-retest hold out sample. We also present estimates of temporal validity based on application of these models to a second sample from the same calendar year, and estimates of external validity based on application of these methods to a separate and subsequent media diary national sample. Finally, we will discuss how the results of this research can be used to tailor recruitment strategies based on the optimal prediction models.

Cashing in on ABS GOLD? Exploring the Utility of ABS Frame Appended Auxiliary Data for Potential Nonresponse Bias Assessment and Adjustment
Anh Thu Burks, *The Nielsen Company*; Lauren Walton, *The Nielsen Company*; Trent Buskirk, *The Nielsen Company*; Michael W. Link, *The Nielsen Company*

Address based sampling (ABS) is a viable sampling methodology due to its near universal coverage of residential households with latest numbers placing coverage at 95% of households (Link and Lai 2011; AAPOR Cell Phone Task Force, 2010). The frame itself provides an alternative sampling solution for coverage issues related to cell phone only homes and hard to reach demographic subgroups (i.e., 18 to 34 year olds, blacks and Hispanics) Moreover, ABS frame data are rich and provide options for stratification, oversampling and nonresponse adjustments that extend way beyond what is available for RDD sampling designs. In this paper we present results from a mixed-mode sample survey from an ABS frame that employed vigorous nonresponse follow-up protocols. All randomly selected households were mailed a survey and a subset of nonresponding households received a follow-up in-person survey attempting to gain participation. Here we assess nonresponse biases for both a continuous measure of media consumption and a binary measure of media access by comparing responses on these outcomes between responding and nonresponding households. We will explore characteristics of responding and nonresponding households that are based on both standard survey household demographic variables as well as ABS auxiliary variables that are measured at the block group. We will further assess the degree to which these variables are related to the survey outcomes and determine the degree to which nonresponse biases can be mitigated using propensity models based on a combination of survey demographic and ABS frame variables. Specifically we will assess the utility of ABS frame auxiliary variables in mitigating nonresponse biases by comparing nonresponse adjusted estimates based on both logistic and random forest propensity models derived using only collected survey demographics as well as those based on both survey demographic and ABS frame variables.

Saturday, May 18
8:00 a.m. – 9:30 p.m.
AAPOR Concurrent Session G

Advances in the Use of Paradata

A Glimpse Inside the Mind of a Respondent: Using Paradata to Improve Online Surveys

Travis Pape, *U.S. Census Bureau*

Traditional quality measures of survey instruments include item nonresponse and survey completion time. In interviewer-administered modes, quality measures sometimes include interviewer observations of respondent utterances or facial expressions. These results are often subjective and cannot describe the reasons behind respondents answer choices or their experiences with the survey. Use of paradata from Internet instruments allows us to get an objective view of the entire survey experience from initial login to final submission. As part of the 2012 National Census Test, the Census Bureau captured paradata from every page of the online instrument, along with respondent answers. These paradata provide rich data related to respondent interaction with the census Internet questionnaire such as break-off rates, help link access, answer changes, and completion times. These data help researchers key in on items that are problematic from a user perspective in a way that is not possible with traditional data analyses, such as response rates. Paradata results allow researchers to focus instrument improvement efforts on items that are known to be problematic for a respondent in a very specific way. This paper will use paradata results from the 2012 National Census Test to identify potential issues that can be resolved for future online instruments and to highlight design features that worked well.

Use of Paradata to Evaluate Medical Expenditure Panel Survey Data and Operations

Lisa B. Mirel, *Agency for Healthcare Quality and Research*; Steven R. Machlin, *Agency for Healthcare Quality and Research*

The use of paradata in survey research has become increasingly valuable in recent years to facilitate monitoring of survey operations and improve data quality. Paradata consists of information about the data collection process in a survey, including interviewer observations, interview language, computer generated time variables for questionnaire sections and numerous other variables. One survey that uses paradata to monitor survey operations and explore improvements in data quality is the Medical Expenditure Panel Survey Household Component (MEPS-HC). The MEPS-HC is a complex multi-stage nationally representative sample of the U. S. civilian noninstitutionalized population with an overlapping panel design. Each year a new sample is drawn as a subsample of households that participated in the prior year's National Health Interview Survey (NHIS) (conducted by the National Center for Health Statistics). Data are collected in the MEPS-HC through a series of five CAPI interviews that cumulatively cover a two year period on a variety of health related issues including health conditions, use of medical care services, charges and payments, and access to care. There is a wealth of MEPS-HC paradata associated with the multiple MEPS-HC interviews and additional paradata information can be obtained by linking to the NHIS. Selected paradata are routinely

used to improve non-response adjustments to MEPS-HC survey weights and have been used for a responsive design pilot study. This paper describes an ongoing evaluation of the association between paradata measures and data quality in the MEPS-HC. In particular, the current evaluation uses descriptive statistics and multivariable modeling to evaluate areas of improvement in the collection of reported health care utilization in the MEPS-HC. The results are interpreted in the context of strengths and limitations of using paradata for improving data quality and monitoring survey processes.

Using Audit Trail Data for Interviewer Data Quality Management **Haoyu Gu, *University of Michigan*; Nicole Kirgis, *University of Michigan***

Audit trail data, the record of actions and entries on computers by computer users, have been collected in many studies using Computer-Assisted Personal Interviewing (CAPI). Audit trail data collected during the National Survey of Family Growth (NSFG) include a record of every key stroke and the time spent between key strokes while interviewers conduct CAPI interviews. Using these data, a data quality dashboard was created in order to monitor data quality at the interviewer level. Indicators include the average time spent on survey questions, the frequency of using help screens, recording remarks, checking errors, backing up in the interview, and the frequency of “don’t know” and “refuse” responses. Principle component analysis (PCA) is used to investigate the relationship between the elements of the interview process. Three factors identified from PCA are included in the dashboard. Two examples will be presented in this paper, showing that by using this data monitoring technique, interviewers with quality concerns can be effectively identified, and the change of the performance of problematic interviewers after intervention can be monitored.

Examining Response Time Outliers Through Paradata in Online Panel Surveys **Jinyoung Lee, *University of Nebraska - Lincoln*; Tarek Al Baghal, *University of Nebraska - Lincoln***

As nonresponse rates and costs of traditional data collection modes increase, more people are becoming interested in Web surveys as an alternative. Although there are great concerns about coverage errors in Web surveys, the simultaneous advantages of Web surveys—timeliness, cost-saving, various design options, and applicability to mixed modes—make them attractive survey modes. This study focuses on response time using paradata and survey responses from the Internet component of the Gallup Panel. Usually, response time is highly skewed. For example, while the average total response time for a Gallup Panel survey in June was 295.15 seconds, the maximum total response time was 4561.24 seconds. To handle outliers with very long response times, Yan and Tourangeau (2008) replaced observations beyond the upper one percentile with the ninety-ninth percentile value and observations below the lower one percentile with the first percentile value, respectively. This study, however, focuses on the outliers themselves, especially those with extremely long response times. Outliers are potentially important because they provide cues to identify respondent behavior and response patterns. In a preliminary analysis, cutting outliers with long response times at certain points excluded nearly one-third of the participants who broke off from the analysis. Also, there were significant differences in the percentages of item nonresponse between outliers and non-outliers. Despite their importance, outliers tend to be excluded from the analysis because of their great leverage to the overall results. Instead of discussing the optimal cutoff points for outliers, this study aims to examine the features of outliers in online panel surveys and suggests that outliers with long

response latencies be investigated for researchers to understand respondent behavior and improve data quality. Exploring response time outliers through paradata may show us a novel way to approach various issues concerning Web surveys.

What Can Paradata Tell Us About Multi-Establishment Business Reporting?

Eric B. Fink, *U.S. Census Bureau*

Paradata are increasingly used to understand respondent behavior and survey outcomes. In this paper, we use paradata to examine multi-establishment business reporting patterns for the 2011 Annual Survey of Manufactures. The ASM offers two main reporting options: paper and electronic. All Business enterprises are mailed a form, but are encouraged to report electronically. Electronic reporting occurs via the downloadable reporting software used by multi-establishment businesses called Surveyor. Enterprises that do not respond initially are subject to nonresponse follow-up. The ASM nonresponse follow-up includes up to four subsequent mailings to the initial mailing and, for select enterprises, analyst phone calls. We combine Surveyor, 2007 Economic Census data, and other ASM paradata for our analysis. Based on our findings, we discuss ideas for adapting the survey during data collection to bring down costs while maintaining or improving data quality.

Adaptive Design at the Census Bureau

Adaptive Design at the Census Bureau—A New Way of Doing Business

Peter V. Miller, *U.S. Census Bureau*

The Census Bureau has made a significant investment in adaptive design, a strategy for more efficient management of survey data collection. The Bureau is engaged developing capabilities for employing adaptive design in all of its censuses and surveys. This panel illustrates a range of efforts in progress. First, we provide an overview of the projects directed by the newly formed Center for Adaptive Design, which include research on adaptive design components, IT system design and outreach and education. Then we offer two papers that detail efforts to develop and validate paradata resources essential to putting adaptive design into practice. One paper concerns data quality of contact information recorded by interviewers in a number of Census CAPI surveys. This information is used to measure the level of effort expended in attempting to interview each case and in estimating the propensity of each case to respond. The second paradata paper details developmental work on an instrument that supplements contact data with interviewer observations of household characteristics. This information may refine estimates of response propensity and offer a means to adjust for nonresponse bias for cases that are not interviewed. The fourth paper describes the process of integrating paradata resources and survey response data to create a set of timely survey metrics. We detail how effort and cost information is combined with response propensity and key survey estimates in a single display in near real time to allow survey managers to track survey progress and execute adaptive design interventions. Finally, we illustrate an application of adaptive design interventions in the National Survey of College Graduates. The test involves both continuous monitoring of key survey indicators and mode switching to increase the likelihood of response in a shorter field period.

An Investigation of Quality of the Contact History Instrument **Dawn V. Nelson, *U.S. Census Bureau*; Julia Coombs, *U.S. Census Bureau***

The Contact History Instrument (CHI) is a standalone Blaise application housed in the Census Bureau's computer-assisted personal interview (CAPI) Case Management system. Beginning in January 2004, Field Representatives (FRs) have used the CHI application to record details about contacts and contact attempts on the National Health Interview Survey (NHIS). Today, all ongoing and some periodic Census CAPI surveys have embraced the CHI. Survey managers in the field and at headquarters rely on CHI data for daily monitoring of survey progress and quality control. Researchers have used CHI data for a wide range of analyses including survey cooperation and nonresponse, optimization of field operations, and effectiveness of respondent incentives. Furthermore, the CHI is an important paradata source for the Census Bureau's adaptive design efforts. Given the wide acceptance and use of the CHI and its importance in the Bureau's adaptive design initiative, it is critical that the CHI data be fit for the uses to which it is put. In this paper, we discuss a recent multi-survey evaluation of the CHI in terms of completeness, reliability, and validity. We identify weaknesses and strengths of the CHI data, and describe our planned research efforts for improving CHI data quality. We end with recommendations for others using similar interviewer-created paradata.

Interviewers as Respondents: Assessing the Usefulness of Neighborhood and Sample-unit Interviewer Observations **Rachael Walsh, *U.S. Census Bureau*; Nancy Bates, *U.S. Census Bureau***

Interviewer observations have recently gained attention in the survey methods literature as a way to enhance both the data collection process and the quality of the data. Adaptive survey design can potentially benefit from visual information collected by interviewers to provide contextual data about interviewer assignment areas. Survey managers can use this information to manage cases better through response propensity models. When they are correlated with both response propensity and the survey variables of interest, interviewer observations can reduce nonresponse bias through post-survey adjustments. This paper includes an assessment of interviewer observations and the potential of these observations for use in adaptive survey design. The 2012 Survey of Income and Program Participation-Event History Calendar (SIPP-EHC) field test included interviewer observations of 3,582 sample units collected by 340 interviewers. Observations included 17 different characteristics of the sample unit and surrounding neighborhood. In this paper, we address the following research questions:

- How successful were interviewers in collecting the observations?
- Are observations predictive of final survey outcomes?
- Are observations correlated with key survey estimates like employment, participation in social welfare and social insurance programs, health insurance coverage, and poverty?
- Does usefulness of observations vary by neighborhood versus sample-unit level observations?
- Do observations have added value beyond the usual contact history data (e.g. doorstep concerns, number of attempts, mode of attempts)?

Developing Survey Metrics for Adaptive Design

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

Adaptive survey design is based on interventions during data collection to achieve strategic survey goals. Intervening in data collection requires access to metrics that integrate paradata and response data. This paper discusses the development of survey metrics and a dashboard display for the 2013 American Housing Survey conducted by the Census Bureau. We will discuss the decision process to identify the key metrics and the configuration of a dashboard to display them in near real-time. This work involves consultation with the Census survey manager and the sponsor to determine which survey response variables to track daily. It entails tracking case completion and response rate to measure survey progress. It also involves the construction of effort and cost metrics to assess continuously the expense associated with progress. Finally, the construction of survey metrics includes measuring the propensity of open cases to respond to further contact attempts. The combination of survey response, case completion, cost and effort and response propensity measures allow the survey manager to adjust field efforts to optimize data quality while containing costs. The dashboard is dependent on an integrated system of paradata and reporting capabilities. Data from several Census Bureau systems (e.g., Field, Payroll) need to be assembled and converted for use as survey metrics. A Unified Tracking System (UTS) in the Bureau has made survey process data from these different systems accessible to a range of survey stakeholders. We discuss the process of refining information provided through the UTS for particular survey requirements.

2013 National Survey of College Graduates: A Practice-Based Investigation of Adaptive Design

John Finamore, *U.S. Census Bureau*

The goals of adaptive design are to attain high-quality survey estimates in less time and at less cost than traditionally executed survey operations. The National Survey of College Graduates (NSCG) will be fielded from February to July of 2013 and will investigate several facets of adaptive design in order to achieve these goals. First, daily processing (editing, imputation, weighting) is operationally expected to reduce the overall time from the beginning of data collection until the final delivery of data and estimates. In addition to operational efficiencies, daily processing will allow the survey team to monitor several quality measures throughout data collection, including R-indicators, benchmarking, stability of estimates, and response propensities by mode. Adaptive design techniques will be directly employed in a mode-switching experiment, where data quality measures will be examined on a weekly basis, and cases will be switched between modes, or put on hold entirely. This experiment is an attempt to allocate resources more efficiently in order to maximize survey quality while minimizing wasted funds and effort. The NSCG uses the American Community Survey (ACS) as its sampling frame and so has a large quantity of data from which to construct propensity models and calculate expected frame totals. For the 2013 NSCG, propensity models calculated using 2010 NSCG data will be applied to 2013 NSCG data for initial locating and response propensity estimation. Those models will be updated with respondent data from 2013 so that adaptive design decisions employ the most up-to-date models available. Daily processing will use respondent data to calculate weighted estimates of frame variables for comparison with expected estimates from the ACS for benchmarking purposes. This talk will discuss the components of adaptive design that NSCG will implement in the 2013 survey, and present examples of data quality measures using 2010 NSCG retrospective data.

Surveying Families and Households

Concordance of Information Collected from Both Members of Low-Income Couples

Daniel J. Friend, *Mathematica Policy Research*; Amber Tomas, *Mathematica Policy Research*; M. Robin Dion, *Mathematica Policy Research*; Debra Wright, *Mathematica Policy Research*; Robert Wood, *Mathematica Policy Research*

Low-income families and couples are often the target of federal policies and programs, particularly social service programs. As part of the evaluations of these programs, researchers collect background data which is used for several important purposes such as 1) describing the characteristics of participants, 2) controlling variables in impact analyses, and 3) analyzing impacts on subgroups. Although data is frequently collected from both members of couples, little is known about how often partners agree on basic household demographics (e.g., income), or how their perspectives on more subjective measures, such as relationship quality, may differ. Although research exists on the level of agreement between proxies and respondents, little research has been conducted on agreement between partners within a couple. Given that analysis may focus on only one partner it is important that we understand how often couples agree or disagree on this basic important information. To shed light on this question, we will analyze data from three studies involving low-income families funded by the Administration for Children and Families, including the Building Strong Families project (a national evaluation of healthy relationship programs involving 4,700 couples), the Couples Decision-Making project (a multi-method study examining decision-making in 46 low-income couples), and the Creating Healthy Relationships project (an evaluation of an intimate partner violence prevention program including 115 couples). We will examine demographic variables (e.g., family structure, income) and relationship variables (e.g., status, quality) and compute a couples' agreement score indicating the degree to which the couples agree on these variables. Additional data sources from these studies (i.e., observational), will be used in regression analyses to explore potential explanations for discordance. Finally, we will discuss the findings implications and applications for future data collection and analysis of families and couples, determining the best respondent.

“S/he Said What!”: The Challenge of Interviewing Both Partners About a Relationship

Jennifer Satorius, *NORC at the University of Chicago*; Colm O’Muircheartaigh, *University of Chicago*; Angela Jaszczak, *NORC at the University of Chicago*; Stephen Smith, *NORC at the University of Chicago*

The National Social Life, Health, and Aging Project is a longitudinal study designed to explore the role of social support and personal relationships in healthy aging. Each wave of multi-mode data collection combines in-home CAPI interviews with the collection of a wide range of biomeasures. Wave 1 was conducted in 2005-2006 with a nationally representative sample of more than 3,000 older adults. Wave 2 was conducted in 2010-2011. To understand from the perspective of both partners the role intimate relationships play in respondents' health, Wave 2 interviewed the cohabitating spouses and romantic partners (partners) of our primary respondents (primes) in addition to interviewing the primes. Given the inclusion of questions regarding health behaviors and relationship quality, we wanted to assess whether the introduction of partner interviews might discourage response from the primes or introduce a bias

in their responses. An experiment was designed in Wave 2 to assess the impact of the change in methodology. Primes were assigned to one of three experimental conditions: 1) Primes were informed in advance of the Wave 2 interview that the partner would be approached for interview; 2) Primes were informed at the end of the Wave 2 interview that the partner would be approached for interview; 3) No request was made for a partner interview. The results of this experiment will inform future decisions on the design of surveys involving data about partnerships. The design permits the assessment of three effects which will be presented at the conference: first, whether the introduction of the partner interviews affects the data from the primes or their response rates; second, whether the timing of the request has a differential impact; and third, whether the responding partners themselves can provide an unbiased estimate of population values.

Validation of Teacher Report as a Methodology for Collecting Information on Student's Cognitive Knowledge and Skills

Kristin Flanagan, *American Institute for Research*; Cameron McPhee, *American Institute for Research*

The Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011), sponsored by the National Center for Education Statistics (NCES) within the U. S. Department of Education, is a nationally representative study of children in kindergarten during the 2010-11 school year. The ECLS-K:2011 will follow these children throughout their elementary school years, culminating data collection in the spring of 2016 when the majority are in fifth grade. During the kindergarten year collection, the ECLS-K:2011 collected information about children's reading, mathematics, and science knowledge and skills both through direct assessment of the child and through teacher report allowing for a unique opportunity to check the validity of teacher-reported data. Over 18,000 children participated in their kindergarten year, from diverse socioeconomic and racial/ethnic backgrounds, in both public and private kindergarten programs. This study will 1) explore the validity of teacher report of children's reading, mathematics, and science knowledge and skills by comparing teacher reports to direct child assessment data, 2) explore the variation in validity by characteristics of teachers and classrooms, such as teacher education; experience; certification; approaches to instruction (e.g., use of whole group versus small group instruction; use of ability groups) and classroom characteristics, such as size and racial/ethnic composition of the class; and 3) explore the variation in validity by characteristics of the child, such as socioeconomic status, gender, and race/ethnicity. Studies of children's growth and development oftentimes rely on methodologies where the teacher provides information on children's knowledge and skills and direct assessment are not included. A study such as this one will provide information to researchers on the validity of teacher report of such information, exploring the possibility of variation of validity by teacher, classroom, and child characteristics. Understanding the sources of validity differences can help researchers interpret survey results as well as design surveys that minimize this variation.

Maintaining Sensitivity to Socio-Cultural Differences in Survey Instruments for Heterogeneous Samples

Rebecca Weiner, *Mathematica Policy Research*

The past several decades have witnessed sweeping changes in the family and left many U.S. children without the support or involvement of their fathers. In response, the federal government created the Responsible Fatherhood (RF) grant programs. To better understand the

effectiveness of such programs, Mathematica Policy Research is assisting the Administration for Children and Families (ACF) with the Parents and Children Together (PACT) Evaluation, a study of a subset of RF federal grantees. African American and Hispanic fathers will comprise a large proportion of PACT's impact study sample. This paper discusses the design of the baseline survey instrument and the use of pretesting techniques to achieve a culturally relevant instrument that is nimble enough to capture the complex family structures of diverse program participants. We drew on several national surveys of similar populations, and consulted with nationally recognized experts in research, practice and policy as we designed the instruments. We conducted cognitive interviews with African American and Hispanic fathers reporting different family configurations, including married fathers and non-residential fathers who had children with multiple partners, to assess question response and sources of response error. Results from the cognitive interviews suggested respondents inaccurately interpreted several key items, particularly those related to men's mental health and wellbeing, and highlighted issues that warranted adjusting the instrument. In response to the cognitive interview results, we modified item sequencing and wording and included a different mental health scale (PHQ-8), which respondents more easily understood in subsequent pretests. We will discuss the pretest process, the findings for the baseline instrument and implications for future survey research with similar populations.

Potential Explanations for the High Net Undercount Rate of Young Children in the U.S. Decennial Census

William P. O'Hare, *U.S. Census Bureau*; Eric Jensen, *U.S. Census Bureau*; Barbara O'Hare, *U.S. Census Bureau*

The Census Bureau's Demographic Analysis (DA) found a net undercount rate of 4.6 percent for children age 0 to 4 in the 2010 Census, higher than any other age group. In addition, the net undercount rate for young children has increased substantially since the 1980 Census. This paper presents three possible explanations as to why young children have a high net undercount rate in the Census and discusses the implications for data collection. One factor which may account for the difference between the DA counts and the Census may be the population estimation technique for children ages 0-4, where net international migration of young children is underestimated. The second set of ideas is related to the Census data collection instrument and processing which may result in under-enumeration of young children. The third category of ideas is related to the households and living arrangements of young children and the extent to which young children are over-represented in hard-to-count places and households. Each proposed cause is described, currently available data are used to assess the ideas, and additional data are proposed to better assess each idea or set of ideas. The undercount of very young children in the U.S. Census has received relatively little attention in the professional literature, yet there are substantial implications beyond the decennial census. For example weighting of survey results often rely on census data. In addition, data collection procedures for capturing accurate counts of very young children in the census can apply to survey data collection. This paper furthers the discussion of this important issue.

Cell Phone Sampling

Improving the Reliability of Survey Items to Assess Telephone Status in RDD Surveys

Vincent E. Welch, *NORC at the University of Chicago*

The reliability and validity of random digit dial (RDD) landline telephone surveying in the United States has been threatened in the past decade by concerns about possible noncoverage bias linked, in part, to a growing number of households giving up their landline telephone and embracing a wireless only lifestyle (AAPOR, 2010). Since the beginning of the last decade, survey researchers have recognized the need to address the mobile phone population in order to ensure full coverage of the population of U.S. households (Blumberg and Luke, 2012). However, the reliability and validity of the items that assess telephone status have not been established (AAPOR, 2010). Over the past year, NORC has conducted a series of qualitative and quantitative research studies aimed at filling in this vital gap in knowledge. Researchers at NORC conducted focus groups and cognitive interviews with dual-phone (i.e., landline and wireless) users to assess the understandability of the current telephone status items in use in many surveys, such as the National Health Interview Survey, California Health Interview Survey, National Immunization Survey, and multiple surveys conducted by Gallup. In-depth probing revealed substantial threats to reliability associated with the wording of telephone status items and response scaling. We found that by altering the wording of the items and the response scaling, we could increase the reliability of responses substantially. Results of a preliminary test of the new scaling will be discussed.

Cell-Phone-Only Voters in 2012 National and State Exit Polls

Michael Mokrzycki, *Mokrzycki Survey Research Solutions*

Courtney Kennedy, *Abt SRBI*

The November 2012 U.S. exit polls included a question on voters' telephone status not only on the national questionnaire—as in elections dating to 2004—but in surveys in 12 states with high rates of early or absentee voting. Nationally and in the aggregate of the 12 states, one-third of voters were cell-only. By state, however, this proportion ranged from 50% in Arizona to 17% in New Jersey; these estimates among voters correlate highly with National Center for Health Statistics modeled state-level estimates of wireless-only incidence for all adults. In many states presidential vote preference differed starkly between cell-only voters (typically more likely to be Obama voters) and those with landlines, but there were exceptions, including the primary and general election battleground New Hampshire. In seven of the 12 states and the national exit poll, there were supplementary dual-frame telephone polls to reach early or absentee voters; cell/landline status and other survey estimates for those respondents will be compared with those for Election Day voters. Characteristics of cell-only and other voters will be compared with data from past national exit polls. Implications for future pre-election surveys and exit polls will be discussed.

The Use of Billing Zip Code and Recent Activity Flags in Cellular Telephone Samples

David Dutwin, *Social Science Research Solutions*

David Malarek, *MSG*

Major sampling companies have recently begun to offer the appending of billing zip codes and recent activity flags to cellular telephone samples. This study considers the utility of these data by investigating first what percent of cell phone telephone numbers receive the billing zip flag, and then, through the use of a large scale national study, by measuring the percent of cell owners whose billing zip actually matches the self-reported zip of their household. Differences by geography, demographics, and characteristics of the zip codes themselves are analyzed to assess the degree of bias inherent in utilizing only sample which has a billing zip flag, and then of respondents who actually qualify for a study by reporting that they in fact live in the zip code(s) targeted by such sample. This paper also considers the distribution of respondents who have a billing zip flag but do not live in their billing zip, and measures the increase in coverage that can be attained by casting a wider net to nearby zip codes, outside of the study target geography. With regard to recent activity flags, we report on the differential telephone dispositions of sample in a large national study by whether each sample record has any recent activity at all; by the prepaid phone flag, and by the recency of use measure. Implications for cellular telephone sampling are considered.

Adjustments for Missing Cell Phone Only Respondents in Repeated Cross-Sectional RDD Surveys

Burton Levine, *RTI International*

Until 2011, the Behavioral Risk Factor Surveillance System (BRFSS), a repeated cross-sectional random digit dial survey, only utilized a landline telephone frame. In 2011, the BRFSS frame was supplemented with a frame of cell phones. To the extent that health behaviors such as smoking are correlated with cell phone only status, the landline-only sample results in noncoverage bias. Trends in health behaviors are confounded with trends in telephone usage and the 2011 sample design change. Specifically, based on BRFSS data, many states saw a downward trend in smoking rates between 2005 and 2010 that reversed in 2011 when the cell phone frame was added. We present methodology to account for the pre-2011 coverage error and the resulting coverage bias. We impute the missing cell phone-only subjects from pre-2011 data with the 2011 cell phone-only respondents. We then reapply the poststratification and recalculate the smoking rates at each time interval. As a result of this procedure the 2005-2010 smoking rates increased, but not uniformly—the later the year, the more coverage error, and therefore the greater the increase in the adjusted smoking prevalence. In some states, before the adjustment, the 2011 smoking rate was the highest for all years between 2005 and 2011; but after the adjustment, 2011 had the lowest smoking rate. This methodology is generalizable to other outcomes that are correlated with cell phone only status in other repeated cross-sectional RDD surveys that added a cell phone component.

Methodological Briefs: Survey Measurement

Improving the Measurement of Big 5 Personality Traits in a Brief Survey Instrument

Matthew DeBell, *Stanford University*; Ted Brader, *University of Michigan*; Simon Jackman, *Stanford University*; Catherine Wilson, *Stanford University*

The 'Big Five' personality traits are the subject of a huge literature in psychology. Part of this literature employs extensive multi-item scales whose length normally precludes their inclusion on representative sample surveys. The Ten Item Personality Inventory (TIPI) has made Big 5 measurement practical in more settings, including representative surveys. However, TIPI's agree-disagree question format invites acquiescence bias. In this paper we report the results of an attempt to improve personality measurement by rewriting the questions to fix the acquiescence problem. We compare the canonical version to an edited version and assess the quality of the resulting data (from a survey conducted by the American National Election Studies in 2012) on several dimensions: completion time, item nonresponse, paired item reliability, and construct validity. We also compare results from both measures in tests of hypotheses about personality's relationship to political attitudes and behavior. We find that completion time and item nonresponse rates are comparable, while reliability and construct validity for the revised TIPI are as good or better than the canonical version by most measures. The results show how better personality data can be obtained at no additional cost by optimizing questionnaire design.

A Comparative Look at Measures of Socioeconomic Status and How Well They Predict Academic Achievement

David Miller, *American Institutes for Research*; Saida Mamedova, *American Institutes for Research*

Socioeconomic status (SES) generally refers to the social standing or class of an individual or group based on economic and social factors. When studies refer to SES levels (low, high, etc.), people may assume that a common definition or measure has been employed. This analysis will examine specific SES measures used across several education studies, including national household surveys, national longitudinal studies, and national and international assessments. Some education studies, such as the Early Childhood Longitudinal Study (ECLS) and Education Longitudinal Study (ELS), produce a composite SES measure based on parents' education, parents' occupation, and household income as reported by a parent of each student. However, in the Program for International Student Assessment (PISA), a composite SES measure is constructed based on student-reported information from the 15-year-olds participating in the study. It is composed of several variables: the International Socio-Economic Index of Occupational Status (ISEI); the highest level of education of the student's parents, converted into years of schooling; the PISA index of family wealth; the PISA index of home educational resources; and the PISA index of possessions related to "classical" culture in the home. In the Trends in International Mathematics and Science Study (TIMSS), 4th-graders' reports of how many books they have at home is often used as a proxy for SES, and the percentage of students in a public school eligible for free or reduced-price lunch has often been used in studies as a proxy for school-level SES. In this analysis, we will describe SES measures used across major national and international education studies. Using regression analyses, we will examine how variation in student achievement within a given study differs if alternative SES

measures are applied. The study aims to better understand the implications of different definitions and measurement of SES, especially as related to student achievement.

Applying “Best Practices” to Questionnaire Design

Darin Harm, *Arbitron*

Arbitron uses a short questionnaire as the first step of a multi-mode data collection process (mailed screener, phone diary placement and mailed diary package) for recruiting the non-landline portion of the population. If a respondent reports being cell phone only or cell phone mainly the household is added to a cell-phone frame and is used to supplement a 2+ list assisted RDD sample. Improving the response rate for the questionnaire is critical to improving the overall response rate of the ABS frame sample since the overall response rate for the ABS sample can never be higher than the return rate to the initial questionnaire. In the summer of 2012, Arbitron redesigned the questionnaire. The goal of this collaboration was to apply best practices in questionnaire design to increase response rates while maintaining data quality. Several modifications were made to the current questionnaire, including making the survey materials more “official”, limiting response modes, and improving visual flow. The redesigned questionnaire will be tested in the winter of 2012. Arbitron’s current questionnaire will be used as the control. Since multiple changes to the questionnaire are being tested simultaneously, it will not be possible to pinpoint the impact of a specific change. However, our goal is to compare the effectiveness of our current questionnaire to the overall effectiveness of a questionnaire that has been redesigned based on “best practices” in questionnaire design. This presentation will examine the impact of the redesigned questionnaire on response rates, data quality, and demographic representation of respondents.

Examining Errors in Medicaid Reporting Across Four National Surveys: ACS, CPS, MEPS, and NHIS

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*

Surveys provide the only source of estimates for the distribution of health insurance in the population, representing a critical source for evaluating the impact of 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. In surveys such as the National Health Interview Survey (NHIS), Medical Expenditure Panel Survey (MEPS) and the Current Population Survey (CPS) Medicaid enrollment counts are always lower than counts available from enrollment data. If enrollees do not report Medicaid, estimates of other coverage or being uninsured will be biased upwards and Medicaid estimates will be biased downwards. If critical questions about the Medicaid undercount are not addressed, public trust (e.g., fiscal and legislative analysts) in health insurance information from surveys will erode and the impact of the ACA will be difficult to evaluate. We extend work from the SNACC team, a multi-phase project examining the Medicaid undercount in federal surveys, to the American Community Survey (ACS). We use linked 2008 ACS data, the first year health insurance variables were available and 2008 monthly Medicaid Statistical Information System (MSIS) data to examine the extent to which Medicaid enrollment is misreported. We compare the magnitude of the undercount and factors associated with misreporting in the ACS to other federal surveys (CPS, MEPS, NHIS). From previous research we know that measuring health coverage is prone to

some level of error and is worse in surveys with extended recall periods; yet bias to uninsurance estimates is minimal. This work provides the first look at the Medicaid undercount in the ACS, a survey that allows us to explore accuracy of Medicaid reporting by survey mode, and is part of a research agenda to further explore patterns of misreporting and the effect on coverage estimates.

Reliability of Parent-Reported Age of Diagnosis for Children with Autism

Stephen J. Blumberg, *National Center for Health Statistics*; Matthew D. Bramlett, *National Center for Health Statistics*; Heather M. Morrison, *NORC at the University of Chicago*; Alicia M. Frasier, *NORC at the University of Chicago*; Michael D. Kogan, *Maternal and Child Health Bureau*

Early identification of autism spectrum disorder (ASD) is an important first step toward making sure that children with ASD and their families are able to access and benefit from early intervention services. Parent surveys could be used to evaluate progress in reducing the age when children with ASD are first diagnosed. Concerns have been raised, however, about parents' ability to accurately recall this information. We used data from two surveys to evaluate the reliability of parent report. Parents of school-aged (6 to 17 years) children with ASD were identified during the 2009-2010 National Survey of Children with Special Health Care Needs, and these parents were recontacted (on average, 9 months later) for the Survey of Pathways to Diagnosis and Services. Both surveys were conducted by the National Center for Health Statistics as part of the State and Local Area Integrated Telephone Survey, and both asked "How old was the child when a doctor or other health care provider first told you that [he/she] had autism or ASD?" The responses across surveys for 1,341 children were highly correlated (Pearson $r = 0.85$) but did not match exactly for nearly half the children (47%). For many (19% of the total), the reported age of diagnosis differed by two years or more between surveys. Differences of this magnitude were more likely for adolescents aged 12 to 17 years (risk ratio = 1.52), for children living in poverty (RR = 1.41), for children whose parents have no more than a high school education (RR = 1.48), and for children ever diagnosed with attention-deficit/hyperactivity disorder, depression, anxiety problems, and/or behavioral/conduct problems (RR = 1.54). Children with 3 or 4 of these emotional or behavioral conditions were more likely to have discrepant parental reports than children with only 1 or 2 conditions (RR = 1.51).

Interpreting Feeling Thermometers Using Demographic Models

Quinn Albaugh, *McGill University*; Stuart Soroka, *McGill University*

Public opinion surveys often rely on feeling thermometers—questions that ask respondents to quantify their feelings towards politicians, political parties, institutions, and social groups on a rating scale (typically from 0 to 100). In theory, these questions provide easily interpretable interval-level scores for survey researchers, but a number of studies suggest systematic differences in the ways in which respondents assign scores. Differing levels and variances in thermometer ratings then make it extremely difficult to interpret each respondent's ratings—particularly with regard to whether a rating is positive or negative. A 60 for one respondent is, in short, not the same thing as a 60 for another respondent. This study suggests that we can overcome these obstacles to some extent by developing models that predict respondents' scores (almost entirely ignoring the object being rated) based on their demographic characteristics, and then evaluate ratings based on their deviation from the respondent's predicted values. In short, predicted means provide an estimate of what a neutral point is likely

to be, given a respondent's demographic characteristics; we are thus able to capture and then account for a good degree of heterogeneity across individuals. Analyses are based on millions of thermometer scores gathered in the American National Election Studies and U.S. General Social Surveys over the past three decades. Results suggest a means of interpreting thermometer scores that is significantly different from what is typical in the literature, and point towards significant gains in the value of thermometer scores for wide range of analyses. In sum, the analysis provides a useful approach that can help researchers interpret thermometer scores in light of individual and group differences.

Maximizing the Accuracy of Final Pre-Election Polls Predicting the Outcomes of Races for the U.S. Senate, House of Representatives, Governorships and the Presidency: A Meta-Analysis

Samuel L. Storey, *Stanford University*

With the wide variance in polling data collected from the most recent 2012 election cycle, many observers have wondered what makes one pre-election poll more accurate than another. Polls that ask similar questions sometimes yield widely different results, and as of yet no one has undertaken a quantitative analysis to determine why. This study fills this void by taking a holistic look at every public pre-election poll taken to predict elections for the House of Representatives, Senate, Governorships, and the Presidency in the days preceding the 2008, 2010, and 2012 elections. We conducted a regression analysis of key variables that characterize a poll, specifically its distance from Election Day, methodology, pool identity, partisan affiliation, geographic location, time in field, and sample size. We determined that common impressions do not always hold true for poll accuracy. For example, data from 2008 and 2010 reveal that while companies that use automated polling techniques frequently claim their polls are superior, human interviews actually tend to yield more accurate results. Additionally, larger sample sizes, closer proximity to Election Day, and larger constituencies are all characteristics of more accurate predictive polling. Currently, we are continuing to collate poll results from the 2012 election to affirm these results. Using the study's results, future pollsters will be able to construct wiser and more prudent methodologies that will provide more precise polls, and constituents will be able to distinguish which companies create the most reliable data to make more informed decisions about the status of an election.

How Does This Look Over There?: Two Experiments in Formatting

Carol Cosenza, *Center for Survey Research/UMass Boston*; Stephanie Lloyd, *Center for Survey Research/UMass Boston*; Lee Hargraves, *Center for Survey Research/UMass Boston*

The Consumer Assessment of Healthcare Providers and Systems (CAHPS®) instruments are usually self-administered, which means that HOW the pages look and are formatted can influence data quality. As part of two different field tests, several experiments of alternative formatting were undertaken. The surveys were funded by the Agency for Healthcare Research and Quality. Experiment 1: Formatting a 0-10 scale. This test, conducted in a university-based health system, used the CAHPS® Clinician & Group (CG-CAHPS) Patient Centered Medical Home adult questionnaire. Different formats for the 0-10 provider rating were tested, including a vertical format (CAHPS® standard) and several horizontal formats (altering placement of check boxes, numeric responses, and text anchors). Experiment 2: Skip patterns – placement of check boxes. Sometimes respondents are instructed to skip over questions based on their answers to

screening questions. When respondents make errors at screening questions, valuable data can be lost. When surveys are coded, we sometimes observe that respondents correctly follow the skip instructions, but fail to check any response box in the screening question itself. To test whether placement of the check boxes makes a difference in skip compliance, a small sample (n=500) were sent questionnaires where the check boxes were placed to the right of the response options (and directly before skip instructions). We compare the data from this test group with those who received the standard version (check boxes left of the response options). This experiment, conducted in a Medicaid population, used the CG-CAHPS Adult questionnaire. Both experiments used a standard 3-contact mailing protocol. The analysis plan for Experiment 1 is to compare differences in means and item non-response. For Experiment 2, we compare rates of errors of omission, where respondents skip questions they should answer, and errors of commission, questions that should be skipped are answered.

Public Opinion and Health Policy

Public Opinion and Health Policy at the State Level

**Claudia Deane, *Kaiser Family Foundation*; Bianca DiJulio, *Kaiser Family Foundation*;
Mollyann Brodie, *Kaiser Family Foundation*; Sarah Cho, *Kaiser Family Foundation***

With the re-election of President Barack Obama, threats to repeal the Affordable Care Act (ACA) have quieted and attention has turned to the states as they face many implementation decisions and challenges. States vary widely in their willingness to expand their Medicaid programs and develop exchanges in which the public can shop for health insurance. Public opinion of the law also varies greatly by state and as is true nationally, it is largely entrenched in partisan leanings. The Kaiser Family Foundation has been in the field monthly with an in-depth, national sample survey of Americans' views of the Affordable Care Act, and using this rich base of data we dig deeper into the national results and analyze opinion of the law at the state level. We explore which states have the most positive and negative views of the law and see how this corresponds with their 2012 vote choice, their progress in implementing the ACA and compare it to a number of state level benchmarks of who the law is intended to assist, namely the states shares of those on Medicaid, uninsured, and living in poverty. As demonstrated by national data, opinion is often ideologically based and this paper explores where opinion and state policy come together or diverge.

Re-Examining Self-Interest as a Predictor of Policy Attitudes Towards Public Health Policy

Stephanie Morain, *Harvard University*

To what extent are the policy preferences of Americans shaped by self-interest? A substantial body of empirical scholarship in political science and public opinion suggests that self-interest has minimal explanatory power in explaining public attitudes. Among the most frequently cited exceptions to this thesis are attitudes towards smoking policy, with smoking status repeatedly demonstrated as a significant predictor of support. However, the studies most frequently cited as evidence of this exception are three decades old. Given dramatic changes in tobacco control policy, these prior studies may not accurately reflect current attitudes. Further, the role of self-interest in shaping preferences toward other public health challenges, including obesity, remains woefully underexplored. Using data from a 2011 online survey of a nationally representative sample of 1817 American adults using KnowledgePanel®, Knowledge Networks' (KN) national

probability-based Web panel, I replicate and extend prior inquiries into the influence of self-interest upon respondent views towards public health policy. I begin by examining whether smoking status influences respondent views towards legal strategies to reduce tobacco use. I then examine whether and how self-interest may also influence respondent views towards legal strategies to address obesity. Consistent with earlier studies, I find current smokers are less likely to support tobacco control measures. However, I find that former smokers are also significantly less likely to support such measures. With respect to obesity policy, I find that body weight does not predict support for policies aimed at shaping the food environment, but does predict support for “individually punitive” policies such as insurance premiums for obesity status and restrictions on the use of food stamps for the purchase of “junk foods.” I propose these results complicate prior explanations of self-interest as a driver of policy preferences, and suggest the need to revisit the role of self-interest in attitudes towards public health policy.

Attitudes and Preferences Toward Health Care and Their Symmetry with Health Insurance Coverage and Medical Expenditure Behaviors **Steven B. Cohen, *Agency for Healthcare Research and Quality***

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 socioeconomic profiles that distinguish individuals with coverage from those who are uninsured, attitudes regarding the need for and value of health insurance coverage may also affect coverage decisions. Given the potential for individuals' health care preferences to influence health behaviors, it is important to measure the population's attitudes towards health insurance coverage and to examine the persistence of these attitudes over time. Individual opinions and attitudes towards healthcare may also visibly impact upon decisions associated with the use health care services and associated behavior with respect to medical expenditures. This study provides a detailed investigation of the degree of alignment over time in health care attitudes regarding the need and value of health insurance coverage based on national data from the Medical Expenditure Panel Survey (MEPS) sponsored by the Agency for Healthcare Research and Quality. Attention is also given to the alignment and associations revealed between the degree of concordance in health care preferences and the persistence in individual coverage and expenditure patterns over time. The utility of these preference measures as significant predictors that serve to identify individuals with persistently high levels of medical expenditures over time is also assessed.

Public Opinion on Medicare Reform **Becky Hanna, *Kaiser Family Foundation*; Liz Hamel, *Kaiser Family Foundation*; Sarah Cho, *Kaiser Family Foundation*; Mollyann Brodie, *Kaiser Family Foundation***

With Medicare spending expected to rise as a share of the federal budget and the nation's economy, policymakers are challenged to find ways to reduce the future growth in Medicare spending, while preserving the quality and affordability of care, and assuring fair payments to plans and providers. The Kaiser Family Foundation has been tracking Americans' views on health policy topics, including Medicare and Medicare reform proposals, through monthly, nationally representative surveys. In the context of Medicare policy proposals and ongoing budget discussions, this paper explores the public's views of the Medicare program and their reactions to policy proposals to change the program, such as raising the age of eligibility, means testing, and the premium support plan put forth by former vice presidential candidate Paul Ryan

and others. Special focus is given to seniors, the current beneficiaries of the program, and the partisan divides that often pervade public opinion on health policy. As the nation faces significant fiscal challenges and policymakers explore ways to reduce the national debt, understanding public opinion on Medicare is crucial as budget debates move forward.

The Effect of Question Wording on Preferences for Prenatal Genetic Testing and Abortion

Eleanor Singer, *University of Michigan*; Mick P. Couper, *University of Michigan*

At intervals since 1990, the General Social Survey (GSS) has asked a series of four questions inquiring into knowledge of genetic testing and attitudes toward prenatal testing and abortion, most recently in 2010. Preferences for prenatal testing for genetic defects are relatively stable over this time period, with almost two thirds of respondents expressing a preference for such testing. Preferences for abortion in case of fetal defect, on the other hand, showed a decline, from 41.1% in 1990 and 41.7% in 1996 to 28.7% in 2004 and 31% in 2010. From 1990 through 2010, the questions about prenatal testing and abortion were framed in terms of 'baby'—for example, 'Today, tests are being developed that make it possible to detect serious genetic defects before a baby is born. But so far, it is impossible either to treat or to correct most of them. If you/your partner were pregnant, would you want (her) to have a test to find out if the baby has any serious defects?' After the 2010 results were released, some researchers questioned whether the answers might have been different had the questions been framed in terms of 'fetus' rather than 'baby.' The word 'baby' had been chosen on the assumption that 'fetus' would be less familiar to respondents, and would therefore lead to more Don't Knows and No Answers. But in the current climate, it seemed possible that the word 'fetus' would carry a more abstract, impersonal meaning and therefore lead to more frequent expressions of preferences for prenatal testing and abortion. To resolve this issue and provide guidance for future administration of these questions in the GSS, we designed a question-wording experiment fielded by TESS. The data have been collected and analyzed and we propose to describe the results of the experiment at the 2013 conference.

Who Consents?...Especially When Linkage or Biological Data are Involved

I Think I'll Pass on That...: Analyzing Differences Between Respondents Who Allow and Reject Consent Requests in the 2006 HRS

Bradley Parsell, *NORC at the University of Chicago*

The Health and Retirement Study (HRS) is a longitudinal panel study supported by the National Institute on Aging and the Social Security Administration that surveys a representative sample of more than 26,000 Americans over the age of 50 every two years. The HRS explores the changes in labor force participation and the health transitions that individuals undergo toward the end of their work lives and in the years that follow. Beginning in 2006, some study participants were asked to consent to a series of physical measurements (height, weight, etc.) and biological collections (saliva and blood). Additionally, a subset of the respondents was asked for their social security numbers for record linkage. To varying degrees, respondents decline to participate in these activities or give information. Similar to the notion that

nonresponse introduces potential bias in survey estimates, nonconsent could potentially lead to bias in the data collected. Using the public data files from the 2006 HRS, we compare demographic variables and key survey estimates for respondents who did and did not consent to the various collections and record linkage request. For each of the different consent requests, we find that the populations of respondents who decline consent are significantly different from those who provide consent. The differences between these respondents may indicate that bias was introduced into the data collected through the activities requiring additional respondent consent. Further, we analyze respondent characteristics and survey measures that may influence a respondent's propensity to give their consent to a given request.

Obtaining Administrative Record Linkage Consent by Mail: Impact of a Sensitive Request on Survey Cooperation Rates and Nonresponse Bias
Celeste Stone, *American Institutes for Research*; Harmoni Noel, *American Institutes for Research*; David Weir, *University of Michigan*

With response rates declining (Groves 2011), researchers are turning to administrative records as an alternate method for collecting rich and comprehensive data from study participants, while also reducing respondent burden and survey costs. Such linkage requests typically require obtaining consent from study participants for sensitive, personally identifiable information (PII) (i.e., Social Security number [SSN]) (Sakshaug et al. 2012). For this reason, these studies generally use interviewer-administered modes, where interviewers can build rapport and address respondents' concerns. Mail is an attractive, cheaper alternative mode. However, little is known about the feasibility of using a mail survey to make such linkage requests and collect the required PII (Fulton 2012). This paper reports the findings from a study testing the feasibility of using a mail survey to obtain participants' authorization to release their Social Security Administration (SSA) records for survey research. A subsample of 4,879 Project Talent longitudinal study participants who had not been contacted in 37 to 50 years were randomly assigned to either a questionnaire only condition or an experimental condition that included a simultaneous request to link to their SSA data by signing a form and providing the requested PII (SSN, DOB, name, signature). The SSA condition also included a three-level prepaid incentive experiment (\$2; \$20; no incentive). This paper will 1) evaluate the impact of the consent and PII request on questionnaire cooperation rates, 2) assess the extent to which any negative effects are mitigated by offering incentives, and 3) examine the sample characteristics (sex, race, personality, aptitude) associated with higher propensities to consent to the request for PII. Preliminary results indicate that the SSA request depressed questionnaire cooperation rates by at least 10 percentage points. However even after the lengthy period of noncontact and with no incentive offered, at least 20% of those asked consented to the mail-based SSA linkage request.

Examination of Item- and Unit- Nonresponse in Population-Based Social Surveys That Seek to Collect Biological Marker Samples From Respondents
Michael Lawrence, *GfK Knowledge Networks*; Curtiss Cobb, *GfK Knowledge Networks*

A large and growing number of population-based social surveys desire to collect biological markers (e.g., saliva, dried blood, nails, cheek swabs, skin, hair) to investigate the role of biology in social behaviors and processes. Part of the growth in interest is due to recent methodological developments that have greatly reduced the financial and administration costs of collecting biological markers, including the ability of respondents to provide samples by mail. At the same time, requests for biological markers from respondents heighten concerns over

privacy and may encourage systematic nonresponse (both unit- and item-nonresponse) that can bias results obtained from studies. However, for most studies, determining whether systematic nonresponse occurred is difficult because it is not possible to know anything about those individuals who choose not to participate in a study; other studies have only limited demographic information from which to understand non-response. In this study, we use GfK's probability-based Internet panel, KnowledgePanel®, and its extensive profile information on panelists to examine demographic, attitudinal and behavioral differences among three groups of respondents that were invited to participate in two population-based survey studies of adults (18+) requesting bio-marker samples: 1) individuals that completed the survey and provided a bio-marker sample (full completes); 2) individuals that completed the survey but failed to provide a bio-marker samples (item-nonresponse); and 3) individuals who failed to complete the survey in its entirety (unit-nonresponse). Understanding the characteristic differences between these three groups can be used to correct for nonresponse bias. Initial results find that while education, an interest in politics and participation in community groups correlates positively with consenting to provide a biosample; conservative ideology is negatively correlated with consent. Once consent is obtained, failure to provide a biosample appears to be mostly random and not systematically related to demographics, attitudes or behaviors.

Interviewers' Influence on Consent to the Collection of Biomarkers **Julie Korbmacher, *Max Planck Institute for Social Law and Social Policy***

This paper examines the determinants of consent to the collection of biomarkers in SHARE with special regards to the role of the interviewer administering the collection. The Survey of Health, Ageing and Retirement in Europe (SHARE) expanded measurements of objective health by collecting a battery of innovative biomarkers. As a pilot study a new module was implemented in the fourth wave of the German SHARE Study which included the collection of dried blood spots. For this measurement the ethics review board requires the respondents' written consent. The interviewer plays an important role in the collection of biomarkers: (s)he is not only responsible for explaining the measurements and reassuring respondents, but is also the one conducting all measurements. Especially in the case of dried blood spots a high level of interviewer skills and trust of the respondent in the interviewers' abilities is necessary. As the interviewer plays such a crucial role in the collection, we examine their influence in this work. Information on them can be drawn from the 2011 interviewer survey of the German SHARE interviewers. This PAPI questionnaire was administered during field training and includes information on general attitudes towards surveys as well as some questions on interviewers' attitudes, experiences, and expectations with regard to the collection of biomarkers. Effects of these variables on the consent rates of the interviewers will be investigated. The design of the pilot study also allows for a comparison of respondents from the panel and the refresher sample. Given that consenting to the collection of biomarkers may require a lot of trust in the interviewer, we expect respondents from the panel who had the chance to build up trust during previous SHARE interviews to be more willing to consent than the respondents from the refresher sample.

Placement, Wording, and Interviewers: Identifying Correlates of Consent to Link Survey and Administrative Data

Joseph W. Sakshaug, *Institute for Employment Research*; Valerie Tutz, *Institute for Employment Research*; Frauke Kreuter, *University of Maryland JPSM & IAB*

Data linkage is becoming more important as survey budgets are tightening while at the same time demands for more statistical information are rising. Not all respondents consent to linking their survey answers to administrative records, threatening inferences made from linked data sets. So far, several studies have identified respondent-level attributes that are correlated with the likelihood of providing consent (e.g., age, education), but these factors are outside the control of the survey designer. In the present study three factors that are under the control of the survey designer are evaluated to assess whether they impact respondents' likelihood of linkage consent: 1) the wording of the consent question; 2) the placement of the consent question and; 3) interviewer attributes (e.g., attitudes toward data sharing and consent, experience, expectations). Data from an experiment were used to assess the impact of the first two and data from an interviewer survey that was administered prior to the start of data collection are used to examine the third. The results show that in a telephone setting: 1) indicating time savings in the wording of the consent question had no effect on the consent rate; 2) placement of the consent question at the beginning of the questionnaire achieved a higher consent rate than at the end and; 3) interviewers' who themselves would be willing to consent to data linkage requests were more likely to obtain linkage consent from respondents.

DC-AAPOR Student Paper Award Winner

Descriptive Analysis of Influences on Consent to Administrative Record Linkage
Jenna Fulton, Joint Program in Survey Methodology, *University of Maryland*

Surveys increasingly request respondents' consent to link survey responses with administrative records. Such linked data can enhance the utility of both the survey and administrative data, yet in most cases, this linkage is contingent upon respondents' consent. With evidence of declining consent rates, there is a growing need to understand factors associated with consent to record linkage. This research investigates the relationship between design characteristics of the survey and consent request and consent rates, and draws upon all available consent rates from surveys conducted in the U.S. with consent requests. There are three components to this research. We first assess whether rates of consent to record linkage have declined overall. The second and third objectives of this research overlap: we describe several characteristics of surveys that request consent to record linkage, and examine these characteristics as potential sources of variation in consent rates. We selected attributes of the survey and consent request that vary across surveys in the target population, for which sufficient information was available in the methodological documentation, and for which we predicted an influence on consent rates. This includes survey mode, sponsor and response rate; whether consent is requested orally or in writing, whether the request takes an explicit or opt-out approach, the topic of the records requested, and any personally-identifying information requested to facilitate record linkage. The results of this study suggest that consent rates are declining over time, and that some characteristics of the survey and consent request are associated with variations in consent rates, including survey mode, administrative record topic, personal identifier requested, and whether the consent request takes an explicit or opt-out approach.

Evaluating Address-Based Samples II

Measurement Consequences of Mode Switching in Multi-Mode ABS Surveys: Experiments in Case Flow Design

Jennifer Vanicek, *NORC at the University of Chicago*; Felicia LeClere, *NORC at the University of Chicago*; Ashley Amaya, *NORC at the University of Chicago*; Kari Carris, *NORC at the University of Chicago*

Multi-mode surveys, among other strategies, have recently been offered as a solution to a decreasing willingness to respond to social and economic surveys and as a companion to Address Based Sampling (ABS) as a method for improving population coverage and increasing response rates. The improvement, however, may come at the cost of measurement error introduced by asking questions using different methods. Responsive mode switching is vital to achieving coverage and response rate gains yet it is not clear if starting and ending modes and switching rules impact key survey statistics by introducing additional bias (LeClere, et al., 2012). In this study, we use a case flow experiment designed to assess the efficiency and performance of a mail-first multi-mode ABS design from Phase 4 (November 2011 – September 2012) of CDC's Racial and Ethnic Approaches to Community Health Across the U. S. Risk Factor Survey (REACH U. S.) to disentangle mode effects from population composition. The experiment was fielded in six of the 28 REACH U. S. communities. Selected sample lines were randomly assigned to a phone-first or mail-first condition. An attempt was made to match each sample address to a telephone number. Only cases that were matched to a telephone number in these six communities were retained for the experiment. We will examine the impact of non-response and starting and ending mode on estimates for key health statistics from the six experimental communities to assess whether differences in responsive design choices have an impact on estimation. Initial results from the experiment suggest that starting with a mail first case flow design and completing data collection by mail do interact to generate higher estimates of current smoking rates. Other health related variables, such as BMI, cholesterol and diabetes diagnoses, as well as diet show limited variation by case flow and ending mode.

The Geographic Nature of Responses to a Web Survey: How Respondents and Their Sentiments Are Subject to Spatial Bias in a Survey of Technology Usage **Ned English, *NORC at the University of Chicago*; Lee Fiorio, *NORC at the University of Chicago*; Michael J. Stern, *NORC at the University of Chicago*; Becki Curtis, *NORC at the University of Chicago*; Ipek Bilgen, *NORC at the University of Chicago***

Web surveys have been seen in recent years as a convenient lower-cost alternative to other modes, without the coverage drawbacks of random-digit dial telephone surveys. At question is what degree of coverage error might be inherent to Web surveys, and how the kinds of people who respond to Web surveys may differ from typical respondents to other modes as well as the population at-large and thus risk the possibility of bias. We broaden on the paper by Fiorio et al. (2012) by using geographic information systems (GIS) and geostatistical models to examine the spatial nature of bias in respondents to a Web survey, and the subsequent impact on reported sentiments. Our research shows that there is a distinct and clustered nature to the demographics of Web respondents, as influenced by linguistic isolation, race ethnicity, and percent households below poverty. In addition, we quantify the spatial bias present in questionnaire data in a survey of technology usage patterns by comparing items from Web respondents to the same on the in-person General Social Survey (GSS). Our paper shows that there is a spatial clustering inherent to Web respondents that is not present to the same degree

in other modes, which implies a new category of coverage bias that has not yet been addressed in the literature. Our research is important to survey research in general because it demonstrates the use of spatial modeling to explore and quantify a new and emerging issue in the field, this being bias inherent in the rapidly-emerging Web mode.

Rural Route Where? : An Examination of Coverage Issues Associated with the U.S. Census Bureau's National Address List

Kathleen Kephart, *U.S. Census Bureau*

The U.S. Census Bureau's Master Address File (MAF) is a national address list that is used for numerous surveys, as well as the decennial census. In order to create an address frame for sampling, or an address list for the decennial census, an extract of the MAF is generated using criteria to determine address validity. One year before the 2010 Census, the U.S. Census Bureau conducted one of the largest dependent address listing operations in the world, utilizing an extract of the MAF. As the first major field operation of the 2010 Census, it was important to provide an accurate address inventory for the census enumeration operations. An accurate inventory reduces census costs and lessens the risk of either omissions from the census or an over-count. We will present added-in-error and deleted-in-error rates for later census operations, as well as the initial canvassing. In addition to presenting the results of 2010 Census operations we explore the characteristics and demographics of areas in the U.S. with poor address coverage. The majority of blocks in the U.S. only required validation and no actions by listers (Boies, 2012). Poor coverage is defined by areas that required a large number of added or deleted addresses from the existing inventory. Another component of our research is addresses with ambiguous statuses; for instance records that were deleted by one 2010 Census operation and re-added by a later, or records that were found on the ground and existed on the MAF, but failed to meet the criteria to be included on the MAF extract. In order to focus limited resources on addresses and geographic areas that require field work, research that allows us to identify potential errors is key.

Improving the Efficiency of Address-Based Frames With the No-Stat File

Bonnie E. Shook-Sa, *RTI International*

Address-Based Sampling (ABS) frames are typically based on the Computerized Delivery Sequence (CDS) file, which the United States Postal Service (USPS) makes available through licensing agreements with qualified vendors. Research based on the CDS file has found the coverage of ABS frames for in-person surveys to be sufficient in urban areas but problematic in rural areas. Because of low rural coverage, researchers often resort to hybrid sampling frames based on both ABS and traditional field enumeration (FE). With a hybrid frame, areas where ABS coverage is expected to be sufficient are allocated to ABS while areas where poor ABS coverage is anticipated are allocated to FE. The more areas that are allocated to the ABS portion of the hybrid frame, the greater the cost savings. Since 2009, the USPS has made available the No-Stat file, a supplement to the CDS file that contains approximately 8 million predominately rural addresses not found on the CDS file. Previous research indicates that supplementing the CDS file with the No-Stat file could be a cost-effective strategy for improving rural ABS coverage for in-person surveys (Shook-Sa et al. 2012). Although the overall coverage gains provided by the No-Stat file are modest, No-Stat addresses are clustered in relatively small geographic areas. This clustered aspect of No-Stat addresses means that they could significantly improve ABS coverage in some localized areas. In a hybrid frame design, these

coverage improvements could move areas that otherwise would rely on FE to the ABS portion of the frame, which would lower field costs. This research measures the efficiencies that are gained by including the No-Stat file in a hybrid frame design.

Too Many Older Homes in Your Sample? Disproportionately Sampling AOH 55+ Addresses from an Address Based Sampling Frame to Improve Sample Representation

Lawnzetta Yancey, *The Nielsen Company*; Lukasz Chmura, *The Nielsen Company*; Scott Bell, *The Nielsen Company*

Historically, the Nielsen TV diary sample has over represented older households. To address the representation of younger households, we have oversampled households with age of head of households (AOH) under age 35; however, we still under represented households with AOH 35-49. One of the benefits of the address based sampling frame is the presence of AOH indicators on a portion of the addresses. In particular, the AOH 55+ age indicator has a 92% accuracy rate for the 55+ age group. In an effort to improve the demographic representation of completed diary homes, Nielsen implemented disproportionate sampling of addresses with the AOH 55+ indicator starting with the November 2011 sample. This paper will review the data used to make the decision to exclude addresses with an AOH 55+ indicator from a portion of the diary sample; and, it will show if we achieved the benefits expected within a recent TV diary survey.

**Saturday, May 18
10:00 a.m. – 11:30 p.m.
AAPOR Concurrent Session H**

Survey Mode and Survey Error

Assessments of Survey Accuracy Through a Multi-Modes National Field Experiment

Bo MacInnis, *Stanford University*; Jon A. Krosnick, *Stanford University*

Several mode studies have assessed the accuracy of telephone and Internet surveys of probability samples and Internet surveys of non-probability samples (Yeager et al. 2011; Chang and Krosnick 2010; Pasek and Krosnick 2010), yielding a general finding that probability-sample surveys are more accurate than non-probability sample surveys. This accuracy gap, some claim, however, may be narrowed or closed by recent developments in sampling for non-probability samples. To supplement this literature to account for newer modes and methodologies, we conducted a large scale mode comparison study in 2012 with a number of leading online panels participating. The study involved administering the identical questionnaire via RDD telephone calls to a national sample of cell phones and land lines, via the Internet with multiple probability samples, and via the Internet with multiple non-probability samples of respondents. The questionnaire included measures of a range of political opinions with a focus on climate change. Simultaneous data collection through multiple modes allowed us to explore the similarity of the measurements made using the various methodologies and to assess

whether the methodologies differed in the degree to which they yielded accurate measurements of the American adult population. National benchmarks of known high accuracy were used to assess the accuracy of data collected. We examined differences between data collection modes in terms of the distributions of political opinions, the relations between opinions, and the relations of opinions with demographics. We also investigated whether the data collection streams differed in the extent to which survey satisficing manifested as well as in the magnitudes of question wording and question order effects. We also explored whether statistical weighting improves the accuracy of the various datasets, and whether the response rate affects the effort's accuracy comparing cases that completed the questionnaire early vs. late during the data collection period.

Web Versus Outbound: A Mode Face-Off Following the Presidential Debate **Jenny Marlar, *Gallup***

Unique events, such as a presidential debate or natural disaster, present researchers with an opportunity, perhaps even responsibility, to capture public opinion. Understanding attitudes immediately following these types of events can inform policy or courses of action. However, conducting surveys in a narrow window of time is challenging and costly, especially using traditional outbound methods. This study compares an outbound and Web study and draws conclusions about the costs and benefits of each. Gallup interviewed respondents following the Presidential debate on October 22, 2012, either via outbound or Web. Outbound telephone respondents were recruited from a nightly tracking study prior to the debate and agreed to be called back immediately following the debate. Web respondents were randomly selected from the Gallup panel, a probability based panel of more than 50,000 members who agree to complete several surveys per month. Respondents were notified ahead of time that they would be asked to participate in a survey following the debate. Web respondents were randomly assigned to receive the survey at one of three points in time: as the debate concluded, one day after the debate, or three days after the debate. The results will be used to analyze several research questions. First, are the Web and outbound components significantly different in terms of response rates, respondents' demographics, and overall results, and does weighting effectively minimize any of these differences between modes? Second, does a Web survey appear to be effective for collecting opinions at a specific point in time? Paradata will be used to explore whether respondents complete the survey at the requested time and if users on mobile devices were more compliant. Finally, results from the three time periods will be analyzed to see if opinion changes over time and to evaluate the benefit of conducting surveys under tight time constraints.

Estimating Measurement Effects of Survey Modes From Between and Within Subject Designs **Thomas Klausch, *Utrecht University*; Joop Hox, *Utrecht University*; Barry Schouten, *Statistics Netherlands***

Measurement effects are a major problem in mixed-mode surveys suggesting that the same respondent potentially provides different answers under different modes. Mixed-mode researchers therefore often need to know the average size of measurement effects (AME) for the questions of their interest. The present paper discusses estimation of AME using two different data collection approaches: a between subject and a within subject (repeated measures) design. Real-world data from an experiment with N=8,800 subjects in The

Netherlands are presented. In the 'between design', subjects were randomly allocated to one mode only (Face-to-Face, Telephone, Mail, or Web). In the 'within design' subjects were first allocated as in the 'between design' and subsequently re-approached after some weeks in a reference mode (Face-to-Face) repeating a large number of questions. Unit nonresponse in both designs represents a threat to full randomization and thus to unbiased estimation of the AME, if confounders relate to the selection mechanism into mode conditions and the outcome variable. Statistical adjustment of missing data is a possible solution to this problem, but it is based on assumptions. Adjustment in 'between designs' assumes that the selection mechanism is ignorable given auxiliary variables. This is often contestable in practice, because some important confounders might not be observed. An advantage of 'within designs' is that it is more plausible to ignore the selection mechanism when conditioning on the repeated measurements. Thereby it is not problematic whether time-related changes of outcomes between measurement occasions occur, because these can be controlled using subjects who are allocated to the reference mode on both occasions (i.e., Face-to-Face). However, 'within designs' need to assume that measurements can be taken independently across time. We compare AME estimates from both designs for questions from the Dutch Crime Victimization Survey applying regression adjustment with propensity score strata as covariates or propensity score weighting.

Asking Questions on Sexual Identity, Financial Well-Being, Sleep, and HIV Testing in the National Health Interview Survey: Exploring Mode Effects
Adena Galinsky, National Center for Health Statistics; James Dahlhamer, National Center for Health Statistics; Sarah Joestl, National Center for Health Statistics; Marcie Cynamon, National Center for Health Statistics; Jennifer Madans, National Center for Health Statistics; Virginia Cain, National Center for Health Statistics

In recent decades research has demonstrated that audio computer-assisted self-interviewing (ACASI) yields greater reporting of socially undesirable behaviors compared to paper-and-pencil questionnaires and various other forms of computerized interviewing (e.g., computer assisted personal interviewing (CAPI)). The bulk of this research, however, has focused on risky sexual behaviors, sexual abuse, and drug and alcohol use. Less is known about mode effects with potentially sensitive topics such as sexual identity or sexual orientation. Over the past year, three field tests were conducted with the National Health Interview Survey (NHIS), a face-to-face, household health survey, to assess the feasibility of 1) asking questions on sexual identity, and 2) administering these and other potentially sensitive items in ACASI. This paper utilizes data collected during the third field test to explore possible mode effects on estimates of sexual minority status, financial well-being, sleep, and HIV testing. The third field test included a split-ballot experiment in which 3,215 adults were assigned to receive the questions using ACASI and 2,237 to receive them using CAPI. Preliminary results revealed no significant differences in prevalence estimates of sexual minority status by mode of administration, while estimates of HIV testing were higher using ACASI than using CAPI. In addition, preliminary estimates of the average hours of sleep in a 24-hour period revealed a shift toward shorter sleep durations in ACASI. Where significant bivariate results emerged, we attempted to diminish or eliminate mode effects in a series of multivariate analyses, controlling for sociodemographic characteristics such as age, sex, race/ethnicity, and education. We discuss the implications of our results for mode choices when administering questions on sexual identity and mental health, and for prior NHIS CAPI-based estimates of sleep and HIV testing.

Changing of the Guard: Effects of Different Self-Administered Survey Modes on Sensitive Questions

Frances M. Barlas, *ICF International*; Wm. B. Higgins, *ICF International*; Jacqueline Pflieger, *ICF International*; Randall K. Thomas, *GfK Knowledge Networks*; Diana Jeffery, *Department of Defense*; Mark Mattiko, *U.S. Coast Guard*

Compared to self-administered questionnaires, socially desirable responses are more likely found with interviewer-administered questionnaires. However, less is known about differences in social desirability bias between different modes of self-administration. This study compared the results for sensitive questions when asked on a paper-pencil questionnaire versus in a Web-based survey. Personnel at selected military installations were randomly assigned to either the paper-pencil or the Web administration. The paper-pencil survey was administered in a group setting, with an interviewer present to distribute and collect the surveys while the online survey was individually-administered at respondents' convenience. All respondents, regardless of mode, were assured anonymity. The surveys were conducted as part of the Health Related Behaviors Survey of Military Personnel, conducted every three years by the Department of Defense and the United States Coast Guard. The largest survey on service members' behavioral health, it asks about a number of activities that can have serious consequences for military careers such as substance use and mental health indicators, as well as a number of highly sensitive topics, including for the first time Coast Guard members' sexuality. Overall, the paper-pencil survey showed fewer drops offs. After controlling for demographic differences and differences in Internet accessibility and use, in the online survey we found lower prevalence estimates of unhealthy or illicit activities, such as heavy drinking or drinking and driving, and higher estimates of socially desirable attitudes and behaviors, such as exercise and safety, compared to the group-administered, paper-pencil surveys. Contrary to the hypothesis that the online administration would be associated with greater reports of undesirable behaviors, we consider the possibility that respondents to the online survey had concerns about anonymity.

Quality of Measurement

Building an Archive of Reliability of Survey Questions

Duane Alwin, *Pennsylvania State University*

This paper presents a design for a public archive of measures of data quality for the typical kinds of information gathering approaches used in survey research. A progress report is presented concerning an ongoing project that is focused on developing a data base of estimates of the reliability of survey measures. Based on nearly 900 individual measures from several large panel survey data sets based on representative samples of the U.S. population, including measures from the National Election Studies, the General Social Surveys, the Health and Retirement Study, and others, this paper reports on the success of the development of a data base for common survey questions implemented in actual surveys. The paper discusses problems in creating a data base containing estimates of question-specific reliability, along with detailed coding of attributes of the questions (e.g. content, response formats, question length, etc.), which can be used to evaluate the optimal properties of survey questions with respect to levels of measurement error. The approach advanced can be used within a meta-analytic framework for assessing the relative quality of measures, and can be used to improve the quality of inferences from survey data. Preliminary evidence is presented from this data base regarding patterns of variation in levels of measurement error linked to concerning survey content, source of information, survey context, and attributes of questions (question form,

number of response categories, labeling response options, explicit Don't Know options, and question length) as a way of demonstrating the utility of the approach. Given that survey measurement is a key ingredient in the majority of social science research, the broader impact of the present project lies in its contribution to the uses of virtually all types of survey data, which can be evaluated in terms of the results of this study.

Can We Have Confidence in Consumer Confidence? Assessing the Temporal Comparability of the Consumer Sentiment Index

Dmitriy Poznyak, *Mathematica Policy Research*; George F. Bishop, *University of Cincinnati*

Along with the Conference Board's Consumer Confidence Index, the University of Michigan's Index of Consumer Sentiment (ICS) has long been regarded as a reliable and valid measure of public opinion on economic conditions in the country. Not only is it considered an essential forecasting tool; the outcomes it produces are a vital force in the movement of U.S. and global stock markets. ICS has also become a central variable in explanatory models of political attitudes and behavior, particularly in time-series models of presidential approval. Given the importance of this subjective indicator, it is surprising that its psychometric properties—particularly its temporal comparability—have not been established. In the absence of such assessment, it cannot be determined whether temporal change observed in consumer confidence is due to a true change in the construct or to methodological changes in its latent factor structure—thus a measurement artifact. Using multigroup confirmatory factor analysis we decompose the Index to analyze the pattern of survey responses to the five questions—each with a 12-month horizon—used to measure Consumer Confidence since 1972. The results confirm that the ICS has the same overall temporal factorial structure. However, only partial equivalence can be established for the Index, indicating that the measurement error associated with repeated measurements over-time is not random. We demonstrate that the meaning-and-interpretation of some of the items—especially personal economic evaluations—varies significantly over time. At the same time, respondents' sociotropic evaluations of the economy remain temporally invariant. Further analysis of trends in response patterns to the personal economic items shows that respondents systematically interpret them in conceptually different ways in times of crisis vs. economic stability. Our analysis raises questions about the temporal comparability of the ICS and suggests that its partial measurement equivalence must be taken into account in deciding whether we can have confidence in consumer confidence.

A Versatile Tool? Applying the Cross-National Error Source Typology (CNEST) to Triangulated Pre-Test Data

Rory Fitzgerald, *City University London*; Lizzy Gatrell, *City University London*; Yvette Prestage, *City University London*

There are certain error sources that are unique to measurement via cross-national questionnaires, or which occur less frequently in single nation studies. Tools that help to identify these errors assist the cross-national survey researcher in producing a higher quality questionnaire in the source language and also facilitate translation. This paper evaluates the Cross-national Error Source Typology (CNEST), which was developed as a tool for improving the effectiveness of cross-national questionnaire design (Fitzgerald et al., 2009). The CNEST has already proved useful when applied to cognitive interviewing data (Fitzgerald et al., 2011). This paper assesses the consistency and versatility of the tool by applying it to triangulated cross-national pre-test data of a module on 'understandings and evaluations of democracy' from

Round 6 of the European Social Survey (ESS). Quantitative data from a face-to-face pilot in Russia and the UK are triangulated with qualitative feedback from interviewers and respondent debriefs in both countries. The CNEST is applied to these pre-test findings to identify and categorise sources of error in the questions, and to develop improved questions or drop a concept from the module where appropriate. This paper highlights the benefits and challenges that accompany the use of multiple pre-testing tools simultaneously.

Does End-User Experience With Government Reforms Diffuse to General Public Opinion? Two Parallel Quasi-Experiments in Colombia

Clifford Zinnes, *NORC at the University of Chicago*; Christopher Nicoletti, *NORC at the University of Chicago*

This study conducts a quasi-experimental impact evaluation to examine several questions regarding the effect of providing citizens an office of transparency and accountability (e.g., accessing freedom-of-information services, lodging complaints of corruption) on public opinion in Latin America on democracy and governance in general and on the quality of government in particular. First, is there an influence on the opinions of end-users of this service regarding their confidence in democracy and governance and does the answer depend on socio-economic or demographic characteristics? Second, how quickly and to what extent is there diffusion of the opinions of direct users of the service to public opinion in general? Are the former's opinions good predictors of the latter's eventual opinions? Third, how does one quantitatively measure such opinions? For this purpose a household-panel public-opinion survey and two end-user cross-sectional exit surveys are administered in treated and comparison municipalities in Colombia at a common baseline in 2010 and endline in 2012. Then multiple applications of propensity-score matching were carried out and difference-in-differences impacts and attribution equations estimated. A novel quantitative indicator design approach is developed and tested to capture inherently qualitative opinions. Among the findings include strong positive improvements as a result of the new office in the opinions of direct users on a range of beliefs concerning democracy and the acceptable ways of exercising it, but slow diffusion of these changed opinions to the general public over the evaluation period. These results, therefore, may serve as a warning of the limited utility of conducting household public-opinion surveys in the short term when gauging the effect of government reform.

Informed Computerized Adaptive Testing: Using Prior Knowledge to Improve Dynamic Surveys

Josh Cutler, *Duke University*; Jacob M. Montgomery, *Washington University in St. Louis*

Survey researchers avoid using large multi-item scales to measure latent traits due to both the financial costs and the risk of driving up non-response rates. Typically, investigators select a subset of available scale items rather than asking the full battery. Reduced batteries, however, can sharply reduce measurement precision and introduce bias. In this paper, we evaluate how computerized adaptive testing (CAT) within a Bayesian framework can (a) minimize the number of questions each respondent must answer as well as (b) seamlessly incorporate prior knowledge about respondents into the survey procedure all while maximizing measurement precision and accuracy. CAT algorithms respond to individuals' previous answers to select subsequent questions that most efficiently reveal respondents' position on a latent dimension. Latent traits of interest may include individuals' political knowledge, healthy eating habits, or propensity to vote. Utilizing information gleaned from prior responses on a respondent's

questionnaire or on a previous panel wave, we demonstrate how, through informative priors, we can increase measurement relative to alternative methods. Using simulations, convenience samples, and a national probability sample, we demonstrate the advantages of using prior information in a CAT algorithm by testing multiple priors and showing that, in most cases, we can achieve greater accuracy and precision when compared to a static battery or a naïve CAT algorithm. We demonstrate how this approach can be used as a dynamic and theoretically motivated way to reduce the size of commonly used batteries (e.g., the big five and need for cognition inventories). We conclude by noting how this method could be extended to include information about respondents gleaned from public records such as voter files. This may facilitate the use of shorter questionnaires to achieve the same levels of measurement quality in a wide array of domains.

Unlocking the Potential of Conjoint Analysis/Discrete Choice Modeling and MaxDiff Scaling in Public Opinion and Survey Research

Motivating Consumers to Participate in Wellness Programs

Lisa Weber-Raley, *Mathew Greenwald & Associates*

Health care policymakers are always seeking ways to improve the quality of health care in this country, while keeping it affordable and accessible. One important strategy is to encourage healthy lifestyles that will help to prevent and/or manage chronic conditions through a wide range of initiatives, loosely termed wellness programs. Today, many wellness programs are sponsored by large employers, but other types of community organizations and health care institutions also offer these types of support services. National health care reform legislation also has promised funding for small employers to offer wellness programs. One key challenge with wellness programs is motivating consumers to participate, so they can have support to maintain or improve their health status, and therefore, lower health care costs. Many large employers, the current leaders in implementing wellness programs, try to encourage participation among their employees by offering incentives. However, there is limited knowledge about what wellness program features effectively motivate participation, what amount or type of incentive spurs participation, or whether features and incentives work the same across different types of wellness programs. Much of the existing research on motivating consumers to participate in wellness programs is case-study based. Understanding the “return on investment” in wellness programs on a broader scale is crucial in order for more public health agencies to justify these offerings. We used discrete choice modeling to study three specific types of wellness programs to identify the optimal feature design to make them appealing to the audiences they target. The programs tested were: Biometric Screening, Exercise, and Health Coaching programs. We conducted online survey of 1,200 employed Americans ages 21 to 65, where each respondent was asked about one of the three programs. Respondents who were selected for the Health Coaching program had to have a chronic health condition or a BMI of 30 or higher.

Message Testing in an Environmental Context **Barry T. Radler, *University of Wisconsin-Madison***

Aquatic Invasive Species (AIS) can cause significant ecological and economic harm to lakes and other water bodies. One of the primary ways AIS spread is by 'hitching' rides with anglers, boaters and other recreational enthusiasts. Although the behaviors these water users need to adopt to prevent AIS "hitchhiking" are fairly simple, behavior change cannot be achieved without carefully planned communication efforts. Various communication strategies have been implemented to increase awareness of AIS problem and encourage behavior change. Many of these strategies have been successful; however, it is not clear what components of current AIS-prevention campaign efforts are having the most impact. Also additional communication efforts are needed to influence individuals who are still not practicing AIS-prevention behaviors. Although social marketing and behavioral theories are frequently used in health communications, little research has applied these theories within an environmental context. As AIS spread is directly linked with specific behaviors, this project presents a unique opportunity to test the effectiveness of key concepts from social and behavioral theories. It is important to evaluate prototypes of different creative strategies to determine if campaign materials are optimally designed to influence attitudes and subsequent behavioral intentions among an identified target audience. We used an online survey that exposed 1,000 individuals from Wisconsin who boated, fished, or recreated on a body of water in the last year to a number of distinct stimuli. Using a discrete choice task with a split-sample design, half the respondents (randomly selected) completed a choice task of AIS materials, while the other half (holdout sample) evaluated the AIS materials using traditional Likert-type response scaling. The holdout sample served as an external validity check of the conjoint model. Multiple dependent measures were employed and the survey also contained behavioral, attitudinal, and knowledge measures regarding AIS that could be used for segmentation analyses.

To Complete by Smartphone or by Tablet or by Computer or by Paper & Pencil – That is the Question: Exploring Factors Associated with Respondent Mode Choice for Multi-Mode Surveys **Trent D. Buskirk, *The Nielsen Company***

Today more than ever before researchers have an unprecedented opportunity to administer surveys via a vast collection of Internet-enabled devices including smartphones, tablets/e-readers, netbooks and desktop and laptop computers. Currently in the U.S., smartphones account for nearly 50% of all cell phones and roughly 20% of households own some type of tablet device. As these penetration levels rise, survey researchers have more viable modes for online survey administration and respondents have more choices with which to complete online surveys. Currently, there is relatively little published research comparing response rates and potential mode effects for surveys completed via these new modes. Even more elusive is the literature that explores the relationship between survey factors, such as recruitment methodology and questionnaire content, and the respondent's choice of completion mode. In this paper we present the results of a conjoint analysis administered to a national probability sample of 1000 smartphone, tablet and personal computer owners aimed at investigating the relationship between five survey attributes (i.e. topic, sponsor, length, incentive amount and delivery type) and a respondent's choice of completion mode (i.e. smartphone, tablet, personal computer and paper and pencil). We will also investigate whether various technology-related variables (i.e. Internet usage and prior survey experience by device) as well as demographic variables (e.g. age, race, education) might explain the latent structure of the derived mode

choice utilities. Finally, we present the results of an experiment randomizing half of the respondents to complete the entire conjoint exercise and the remaining half to complete a subset of the conjoint questions. We present external validity measures derived by comparing the modeled preferences of respondents assigned to the full conjoint exercise to the observed preferences from those assigned only to the subset questions.

Price and Preference Sizing for a Consumer Service **Mario Callegaro, Google UK**

We will field a Choice Based Conjoint survey for an online consumer service. Goals of the project are to establish: 1) Interest and tradeoffs among performance features of the service, 2) Brand value, 3) Price sensitivity, 4) Preference share vs. competing services. Additionally, this project will address questions of conjoint analysis replicability, internal reliability, and external validity by comparing results to a previous sample, to current market share, and to a real, inline indicator of interest presented at the end of the survey. The sample will comprise approximately N=1650 adult, online, general consumer respondents in the U.S. We will include approximately K=20 demographic, attitudinal, and behavioral survey items in addition to the conjoint; total survey time of 10-15 minutes. Previous work suggests that conjoint analysis is internally reliable but may need per-project assessment of external validity (Chapman et al., 2009*). We will compare the results here to: 1) Results from a previous CBC study that assessed a partially overlapping set of attributes/levels, 2) Actual current market share in both a regional area and national area. We believe these comparisons will be of particular interest in the survey analyst community.*CN Chapman, J Alford, and E Love (2009). Exploring the Reliability and Validity of Conjoint Analysis Studies. Presented at the 2009 Advanced Research Techniques Forum, Whistler, BC, June 2009.

State of the Art: Past, Present and Future of the Survey Profession

Old and New Survey-Research Paradigms **Tom W. Smith, NORC at the University of Chicago**

A paradigm shift occurred almost 80 years ago in the mid-1930s when Gallup, Roper, Crossley, and a handful of other innovators pioneered the public opinion poll (Brick, 2011; Converse, 1987; Groves, 2011b). Prior to the advent of polling, politicians, journalists, social scientists and others had turned to various sources to measure public opinion and other aspects of society. These included tracking election returns, the outcomes of referenda, crowd counts, straw polls, compilations of editorials and news articles collected by such publications as Public Opinion (taken over by Literary Digest in 1906), studies of letters to the editor, and, as George Gallup (1957) noted in 1957, such other evidence as “letters to congressmen, the lobbying of pressure groups, and the reports of political henchmen...” These alternatives were supplanted by the polls and soon public opinion and poll results became considered to be almost synonymous with one another. The advent of polling was a complete game changer. As Elmo Wilson (1947), a researcher at Roper and other organizations, remarked in 1947, “25 years ago the possibility of measuring public opinion with any degree of precision was at least as remote from public consciousness as the atomic bomb.” Now a rising chorus is asserting that polls are passé, a growingly antiquated relic of the last century. They claim that public opinion, consumer behaviors, and other socio-political outcomes can be better measured (less expensively, more

quickly, more easily) by the analysis of Internet usage in general and of social media in particular, by the data mining of administrative databases (including the merging of disparate information sources through such techniques as data fusion), or by a combination of these two alternatives to traditional surveys. The promise and pitfalls of this new proposed paradigm are considered.

The Evolution of Presidential Polling

Robert M. Eisinger, *Savannah College of Art and Design*

Interest in presidential polling continues to grow. What are they? How are they conducted? Why and when? 2013 will mark the 10 year anniversary of the publication of *The Evolution of Presidential Polling* (Cambridge U. Press). The 2012 election and related media coverage underscores the interest in polls, and the continued discussion between presidential leadership and responsiveness to public opinion. This proposed panel explores the past, present and future of presidential polling, with the goal of educating attendees and exploring new theories about how polls are conducted and why.

Self-Reported Participation in Research Practices Among Survey Methodology Researchers

Kelly Perez-Vergara, *Independent Consultant*; Caroline Smith, *Dana-Farber Cancer Institute*; Carol Lowenstein, *Dana-Farber Cancer Institute*; Al Ozonoff, *Boston Children's Hospital*; Yolanda Martins, *Boston Children's Hospital*

In recent years, the issues of accountability, transparency and ethical conduct in scientific research have received widespread media attention. However, the ethical "grey-zone" of research practices is widely exploited, as demonstrated in John et al.'s (2012) report that more than 63% of investigators admitted they had failed to report all of a study's dependent measures in a published paper and over 45% admitted that in a paper they had 'selectively reported studies that 'worked'.' We do not know of any published studies in the area of survey methodology that quantify how often researchers employ various research methodologies or the implications of their use, when conducting research about survey methods. In order to assess the use of and beliefs about various research methodologies and practices that may be utilized while conducting survey methodology research, 483 men and women, identified through systematic Web searches as survey researchers, were invited to participate in a Web survey. The survey included 14 items assessing demographic variables, 10 items related to use of and belief about methodological designs used in survey methodology research and 15 items on beliefs about ethical conduct of research. Results will be discussed in terms of the potential ethical implications and the American Association for Public Opinion Research's commitment to transparent survey research methodologies.

Transparency in the 2012 Pre-Election Polls

Stephanie Calvano, *Marist Institute for Public Opinion*; Daniela Charter, *Marist Institute for Public Opinion*; Michael Conte, *Marist Institute for Public Opinion*; Natalie Jackson, *Marist Institute for Public Opinion*; Susan McCulloch, *Marist Institute for Public Opinion*

Are pollsters providing enough information? In 2009, AAPOR began the Transparency Initiative, designed to "encourage routine disclosure of methodological information from polls and surveys whose findings are released to the public." The Initiative is applicable to all types of polling and

survey data, but perhaps the highest volume of publicly released survey findings occurs prior to U.S. Presidential elections, making these polls a particular focus of scrutiny. Some of the firms that release pre-election polling numbers are members of AAPOR and have signed on to support the Transparency Initiative, and some are not part of the Transparency Initiative or AAPOR. The variety of firms that produce and release pre-election polls provide an ideal opportunity to evaluate the transparency of various organizations, and how easily their methodological information is accessed as well as what information is provided. In this meta-analysis, we will review the polls reported by Real Clear Politics in the months before the 2012 presidential election, plus any others used by poll aggregating models, and determine two things: 1) how much effort is required to access their methodological information online, and 2) how much information is provided in their methodology statement. In an atmosphere in which pre-election polls are under heavy attack, methodological transparency is of utmost importance. We will report how methodologically transparent public polls were during the general election of the 2012 presidential campaign. Data will be aggregated by type of organization, for example AAPOR members and those who have signed on to the Transparency Initiative vs. non-member organizations, academic vs. non-academic organizations, and partisan vs. non-partisan polling organizations. The research design will include ease of accessibility to information by both experienced researchers and evaluators with little or no research background.

Trust in Statistics and Statistical Use of Administrative Records

A Multi-Method Analysis of Measurement Error Using a Measure of the Public's Trust of Official Statistics in the United States **Morgan Earp, U.S. Bureau of Labor Statistics**

In an effort to explore the public's trust of official statistics in the United States and attitudes towards the use of administrative records, the Census Bureau collaborated with several federal statistical agencies to develop a measure of trust in statistical products, trust in statistical agencies, and attitudes towards use of administrative records. This measure is being used in a telephone survey to monitor the public's trust level and assess the impact on attitudes towards use of administrative records. During the construct refinement and item development phase, we consulted international models of trust of official statistics (Brackfield, 2011; UK Office for National Statistics, 2006 & 2007). Prior to pretesting, cognitive interviews and expert reviews were used to assess and improve items. Pretesting was done in three phases, allowing us time to assess and address measurement error between administrations. During pretesting, we used random probes to assess item performance and we used confirmatory factor analysis (CFA) to evaluate item misfit (error variance) within factors. Since pretesting was completed, we have continued using the prior methods as well as Item Response Theory (IRT) to evaluate items. While the results from each analysis are correlated to varying extents, it appears that each tool taps into a unique aspect of measurement error, and that no single tool provides a complete assessment. While the results from some tools are weakly correlated with item nonresponse, the results from other tools are strongly correlated with item nonresponse. This paper focuses on the relationship between the various diagnostic tools used to assess measurement error and the relationship between measurement error and item nonresponse. We will present the theoretical model we developed, the methods used to detect measurement error, and our analysis of the relationship between item nonresponse and measurement error.

Monitoring and Detecting Shocks that Influence Change in Public Trust Towards the Federal Statistical System

Melissa A. Mitchell, *USDA/NASS*

Beginning in 2011, several federal statistical agencies partnered to develop a measure of trust in official statistics and monitor public opinion on the use of administrative data for statistical purposes. Using the Fellegi model of trust of official statistics as a starting point, Earp and colleagues (2011) identified factors related to trust and public perception of the Federal Statistical System. These factors are: trust in statistical products (accuracy, relevance, and credibility), trust in statistical institutions (integrity, confidentiality, transparency, and impartiality), and *trust* in official statistics. It is hypothesized that these factors may influence attitudes towards the use of administrative records for statistical purposes. Considering trust and perception can change over time and could be influenced by many different, external events, we planned to study these factors over time to see if trust and perceptions towards the statistical system and opinions towards the use of administrative records are changing, and, if so, what influences their change. Using time series techniques, we examine these factors over time. We are interested in external events that may occur that cause a “shock” to the system. A shock is defined by its location in time and its magnitude. It can have both an immediate impact as well as a long-lasting impact. Shocks are reflected by the residuals (error terms) once an adequate model is fit. Part of this study is hypothesis driven, for instance, events that make an impact in the media, like former CEO of General Electric, Jack Welch, questioning the validity of the unemployment rate, and the presidential election may impact trust and perception. In addition to hypothesis driven inspection, we also employ a retrospective approach where we look for changes in opinion and see if we can determine events in the media that may have coincided with the change in opinion.

To Share or Not to Share? Understanding Respondents’ Privacy and Confidentiality Concerns Regarding Administrative Records Usage

Michelle Smirnova, *U.S. Census Bureau*

The U.S. Census Bureau is investigating the use of administrative records, which could create unease if respondents believed that the agency was treating their personal data inappropriately. Although privacy and confidentiality are protected by different laws, the two concepts are often conflated in respondents’ minds. This creates a problem in measuring these concerns and designing effective communication strategies to address them. Accordingly, the Census Bureau collaborated with other agencies to conduct focus groups and cognitive interviews to design a questionnaire that would measure privacy and confidentiality concerns of respondents regarding the use of administrative records for statistical purposes. In a series of three focus groups and 85 cognitive interviews, we explored respondents’ concerns with the use of administrative records data, which allowed us to formulate and refine survey questions that measured the constructs as intended. We found that respondents do not have consistent opinions regarding data sharing; rather their reactions depended largely on two factors. The first was data-specific: if respondents believed that data collected by another agency was accurate, beneficial to society, or cost-effective, they had favorable attitudes. The second factor was agency-specific: respondents tended to divide organizations into two categories—benign information-gathering agencies, whose use of data is perceived to have either a positive or neutral effect on the respondent versus sanctioning agencies whose data use is associated with negative consequences. For this second group of agencies, even if data were perceived as accurate, beneficial or cost effective, respondents were opposed to another agency sharing their personal

information with this perceived-as-threatening organization. This research enabled us to separate privacy and confidentiality concerns, utilizing the results to design more precise survey questions and to craft messages that future public relations communications campaigns could use to allay respondents' concerns about privacy and confidentiality with regard to administrative record usage.

Predicting Attitudes Towards the Use of Administrative Records

Ryan King, *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 are asked at the end of an ongoing nightly telephone survey. The survey is being fielded from January 2012 to September 2013 and completes interviews with about 200 nationally representative respondents most nights. Respondents are asked a number of questions regarding their attitudes towards and knowledge about the Federal Statistical System, as well as questions about their attitudes and knowledge of the potential use of administrative records data for statistical purposes. Building on past research in this area, through the nightly survey, we have examined various ways of measuring and influencing opinions towards the use of administrative records. This paper explores overall attitudes towards administrative records use and compares whether mentioning different social benefits (such as saving money or time), using different data sources (such as government, commercial, or health records), and different federal agencies requesting use of the record may produce different results. 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. We also show how this line of research can be used to help frame the public discussion of the use of administrative records for statistical purposes.

Mixed Topics in Questionnaire Design I

Estimation of Expected Academic Engagement Behaviors: The Use of Vague Quantifiers Versus Talled Responses

James Cole, *Indiana University*; Alex McCormick, *Indiana University*

This study sheds light on a rarely explored topic in survey research: do different behavior estimation procedures for past and expected behaviors produce different results? This study is based on prior research regarding the importance of academic expectations, estimation of behavior frequency (e.g., Schaeffer & Presser, 2003), and the use of vague quantifiers in survey research (e.g., Wright, Gaskell, O'Muircheartaigh, 1994). Data for this study are from the 2010 administration of the Beginning College Survey of Student Engagement. Responses from more than 28,000 first-year students enrolled at 68 institutions were included in this analysis. Items from the core survey were repeated at the end of the Web version of the survey. Respondents were reminded of their original response to the item (core survey items are presented with vague quantifiers: very often, often, sometimes, and never) and were then asked to again estimate their behavior by tallying or counting their behaviors. One of the general findings is that the magnitude (effect size) of the differences for the vague estimations was much larger than for the tallied estimations. This means that those doing "gap analysis" where data are used to identify areas where student expectations are not met, may want to consider if the results are more of an artifact of the response set, than any real difference in behavior frequency. This study also found that tallied estimates associated with vague quantifiers are not necessarily stable. For instance, in high school "very often" asking questions in class corresponded with a tallied count of this activity of 23 times per week. However, "very often" expecting to ask questions in class during their first year of college corresponded with a mean of 16 times per week ($d_{pooled}=.550$). Full results will be presented and implications for survey research discussed.

Numeric Estimation and Response Options: An Examination of the Measurement Properties of Numeric and Vague Quantifier Responses

Tarek Al Baghal, *University of Nebraska - Lincoln*

Many survey questions ask respondents to provide responses that contain quantitative information. These questions are often asked requiring open ended numeric responses, while others have been asked using vague quantifier scales. Generally, survey researchers have argued against the use of vague quantifier scales. However, no study has compared accuracy between vague quantifiers and numeric open ended responses. This study is the first to do so, using a unique data set created through an experiment. 124 participants studied word lists of paired words in the experiment, where the experiment employed a 2 (context: same; different) x 6 (frequency of target word presentation: 0, 2, 4, 8, 12, 16) x 2 (response form: open-ended numeric; vague scale) factorial design, with the context and form factors manipulated between subjects, and the frequency factor manipulated within subjects. There are two conditions for the context factor: same-context condition where the same context word was paired with each presentation of the target word and different-context condition where the same context word was paired with each presentation of the target word. The other between subject factor was response form, where participants responded to a recall test using either vague quantifiers or numeric open ended responses. Translations of vague quantifiers were taken and used in accuracy tests. Finally, a numeracy test was administered to collect information about respondent numeracy. Different accuracy measures were estimated and analyzed including

relative accuracy, bias in estimation, and signed and absolute differences. Results show context memory did not have a significant effect. Numeracy has an effect, but not always in the same direction, depending on form and context. Actual frequency had a significant effect on accuracy, but did not interact with other variables. Importantly, response form does not always have impact on accuracy, but when it does, vague quantifiers tend to improve accuracy.

Including Covariates in a Factor Mixture Model Intended to Detect Differences in Vague Quantifier Interpretation

Jamie L. Griffin, *Mathematica Policy Research*

Survey respondents are commonly asked to provide vaguely quantified estimates of behavioral frequency (e.g., never, sometimes, often, very often). Researchers interested in placing respondents on a latent behavioral frequency continuum based on a set of related items often assume that the interpretation of these vague quantifiers is identical across respondents—for example, that all respondents interpret sometimes as 1 to 2 times or very often as 5 or more times. If this assumption is incorrect, detected differences on a latent factor estimated from these frequency reports (e.g., student engagement) might reflect differences in interpretation rather than true differences on the factor. Several studies investigating the interpretation of vague quantifiers have demonstrated that individual variability is not necessarily random; rather, the variability tends to be associated with demographic or social characteristics (for example, education, age, race, social class). Griffin (2012) described the use of a factor mixture model to detect latent “interpretation” classes; that is, unobserved groups of respondents for whom interpretation is consistent. There was, however, wide variation in the ability of the model to correctly predict vaguely quantified responses; thus, the extracted latent classes may not necessarily differentiate respondents according to their interpretation of vague quantifiers. The present paper investigates whether the model’s performance is improved by including covariates representing social referent groups (e.g., gender, class rank, race). Using data from an experiment embedded in the 2006 National Survey of Student Engagement in which 8,174 students reported frequencies of several student engagement behaviors in both numerically and vaguely quantified terms, we first outline how to include covariates in a factor mixture model estimated on vaguely quantified frequency reports intended to detect differences in the interpretation of vague quantifiers. Second, we evaluate the model’s performance by comparing the numeric frequencies estimated from the model to those directly reported by the students.

Validating Sensitive Questions in Labor Market Surveys: A Comparison of Survey and Register Data

Antje Kirchner, *Institute for Employment Research (IAB)*

The randomized response technique (RRT) is one of the most popular and best investigated so-called “dejeopardizing techniques,” a class of data collection strategies for eliciting sensitive information. This paper explores the RRT as a means to improve the quality of data about sensitive labor market topics, such as receipt of basic income support. In a 2010 telephone survey (n=3,211), we experimentally tested two techniques for asking such sensitive questions: direct questioning and the randomized response technique. First, we compare the percent of socially undesirable responses (indication of transfer payments, i.e. receipt of basic income support) across the two techniques. In addition, because the sampled persons were selected from German administrative records, we know (in the aggregate) the percent of respondents who have received transfer payments and thus the percent who should have reported receipt.

Thus we can also validate the reported percent from each method against the known true rate for the responding cases, hence assessing the bias of our estimates. Such administrative record data is quite rare in the literature on sensitive questions, and allows us a unique opportunity to evaluate the “more is better” assumption which is so often used in the literature. Being able to assess the amount of ‘non-compliance’ to the RRT instructions for this item, insights into the functioning of the RRT also in specific sub-populations can be assessed using multivariate analyses. Thus this paper provides insights into a variety of practical and theoretical factors contributing to successful implementation of the RRT in labor market surveys.

Are Readability Formulas Valid Tools to Assess Survey Question Difficulty?

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

Readability formulas, such as the Flesch Reading Ease formula (Flesch, 1948), the Flesch-Kincaid Grade Level index (Flesch, 1979), and the Gunning Fog index (Gunning, 1952), are often considered to be objective measures of language complexity. Not surprisingly, survey researchers have frequently used readability scores as indicators of question difficulty (e.g., Converse, 1976; Ganassali, 2008; Harmon, 2001; Holbrook et al., 2006) and some have even suggested to apply the formulas during the questionnaire design phase to identify problematic items and to assist survey designers in revising these questions (e.g., Velez & Ashworth, 2007). At the same time, the formulas have faced severe criticism, in particular for being mostly based on only two variables (word length and sentence length) which may not be very good predictors of language difficulty (e.g., Oakland & Lane, 2004). The present study examines whether the three readability formulas presented above reliably identify problematic survey questions. Readability scores were calculated for a large number of question pairs, each including a problematic (e.g., syntactically complex, vague, etc.) and an improved version of the question. The question pairs came from two different sources: (1) existing literature on survey design (e.g., Fowler, 1992; Fowler & Consenza, 2008) and (2) the Q-BANK database (NCHS). The analyses revealed that the readability formulas often favoured the problematic over the improved question version. On average, the success rate of the formulas in identifying the difficult questions was below 50 percent. Reasons for this poor performance as well as implications for the use of readability formulas in survey research are discussed.

Implementing a Responsive Design: Moving From the Theoretical to the Practical

Using Predicted Response Propensities for Bias Reduction

Dan Pratt, *RTI International*; Melissa Cominole, *RTI International*; Jeff Rosen, *RTI International*; Bryan Shepherd, *Abt SRBI*; Peter Siegel, *RTI International*; David Wilson, *University of Delaware*; Jennifer Wine, *RTI International*

How response rates are increased during nonresponse follow-up can affect the amount of nonresponse bias evident in survey estimates. A common approach has been to simply maximize response rates by targeting sample members who are most likely to be interviewed. However, since nonresponse bias is a function of the association between the likelihood to be interviewed (response propensity) and the survey variable of interest, interviewing the easiest cases during nonresponse follow-up may not reduce bias (e.g., Curtin, et al., 2000; Keeter, Miller, Kohut, Groves, and Presser, 2000). In fact, nonresponse bias may actually increase

when nonresponse follow-up efforts target likely respondents (Merkle and Edelman, 2009). This paper reports the results from three national field test studies which tested whether or not encouraging participation among low propensity (low likelihood) cases can be a practical and effective method to improve overall survey estimates. Various sources of information were used to evaluate propensity: paradata from early interview attempts, demographic and substantive survey data from prior survey waves, and administrative data. The likelihood of any sample member becoming a nonrespondent was estimated prior to data collection and, for those sample cases least likely to respond, a different survey protocol was employed to gain cooperation. The approach rested on the assumption that low propensity cases, which are frequently excluded due to nonresponse, were fundamentally different from responding cases, and their inclusion would reduce bias in key survey estimates.

Comparative Evaluation of Metrics for Tracking and Assessing Nonresponse Bias **Peter Siegel, *RTI International*; Bryan Shepherd, *Abt SRBI*; Melissa Cominole, *RTI International***

Recent work regarding survey error has helped clarify the effects of nonresponse on survey estimates. One of the key findings from this new literature is that response rates are not good predictors of bias. In other words, increases in response rates do not necessarily decrease bias in estimates, a finding that stands counter to the prevailing mindset of many survey researchers. Rather, this research illustrates that a key factor in determining the level of nonresponse bias is the covariance between response propensity and response values within the population of interest. In cases where the response value for a survey item varies with propensity to respond to that survey item, nonresponse can create bias in estimates. In cases where this relationship does not exist, or is obscured by other relationships or survey errors, bias due to nonresponse may be absent or hidden. These insights refine our understanding of the roots of nonresponse bias, but in doing so complicate what was once a straightforward recommendation for minimizing nonresponse bias—increase response rates. Fortunately, although still in the early stages, research on new metrics for tracking and assessing nonresponse bias has begun. In this manuscript we consider the impact of monitoring and responding to two specific metrics—Mahalanobis distance and the R-indicator—in the context of a responsive data collection design aimed at reducing nonresponse bias in key survey outcomes. We do this via simulations based on data collected by a large, nationally representative panel survey. We find that these metrics can be useful in gauging potential contributions to nonresponse bias, each with its own pros and cons, and present recommendations for real-world implementations based on the simulations.

Using Mahalanobis Distance Measures for Bias Reduction **Melissa Cominole, *RTI International*; Dan Pratt, *RTI International*; Bryan Shepherd, *Abt SRBI*; Peter Siegel, *RTI International*; David Wilson, *University of Delaware*; Jennifer Wine, *RTI International***

Building upon the results of the experiments and simulation studies discussed in other papers in the panel, our most recent data collections incorporated the Mahalanobis distance measure into a responsive design intended to reduce nonresponse bias among high-distance cases, or those nonrespondents most unlike those who have already responded. We will describe the designs of three studies, each with a unique approach adapted to its population. Each study began with an early response phase during which sample members were invited to complete a self-administered Web interview. After the initial early response phase, outbound telephone prompting and production telephone interviewing began. Each study identified a series of time

points, after the early response and initial outbound calling phases, during which the Mahalanobis distance values were calculated for all remaining nonrespondents, so that cases above a certain threshold could be targeted for specialized protocols. The timing and nature of interventions varied according to the specific needs of each study. The particular design used in each study will be described and preliminary results will be presented. Issues related to practical implementation within constrained budgets and schedules will be discussed.

Using Propensity Models During Data Collection for Responsive Designs: Issues with Estimation

James Wagner, *University of Michigan*; Frost Hubbard, *University of Michigan*

Responsive designs often use response propensities estimated during data collection. These estimated propensities may be used either for monitoring or for making decisions about the next action to take on each case. A problem with estimating response propensities in this way is that the data are not fully observed until the end of the study. The data about future effort and response are “missing.” This missingness may or may not bias the resulting estimates. This presentation reviews situations under which this missingness may lead to bias, discusses approaches to estimation that may minimize the risk of bias, and gives several examples that evaluate the impact of this missingness on estimates and actions taken as a result of these estimates.

Does Balancing Survey Response Reduce Nonresponse Bias?

Barry Schouten, *Statistics Netherlands*

Recently, various indicators have been proposed as indirect measures of nonresponse error in surveys. The indicators employ available auxiliary variables in order to detect nonrepresentative or unbalanced response. They may be used as quality objective functions in responsive and adaptive survey designs. In such designs different population subgroups receive different treatments. The natural question is whether the decrease in nonresponse bias caused by these designs could also be achieved by nonresponse adjustment methods that employ the same auxiliary variables. In this paper, we discuss this important question. We provide theoretical and empirical considerations on the role of both the survey design and nonresponse adjustment methods to make response representative or balanced. The empirical considerations are supported by a wide range of household and business surveys.

Economic Issues and Attitudes

Media, Public Opinion and Economic News Coverage

**Stuart Soroka, *McGill University*; Dominik Stecula, *University of British Columbia*;
Christopher Wlezien, *Temple University***

Public reactions to the economy have political consequences. Support for governments and policies follows economic trends, for instance. But past work shows that media coverage of the economy matters to public attitudes, above and beyond the economy itself; and that coverage is biased, driven by organizational factors, news norms and audience interests. This paper examines one new aspect of the media-public-economy relationship: the tendency for both media and public opinion to react mainly to changes in the economy, conditional on levels. That

is, both media and the public do not react to high unemployment so much as an increase in the rate; and that coverage is conditional on current unemployment levels. This pattern comports nicely with research on voter behavior and elections, which shows that economic change, not the level of the economy, is what matters; it also makes more understandable the somewhat surprising finding of positive coverage in the midst of the Great Recession. Results indicate that the model applies at other times and in other places—they are based on a content analysis of 150,000 news stories (over 20 years) in the U.S., UK and Canada, analyzed alongside commercial polling data on economic sentiment. The paper considers implications of the media-economy relationship for economic sentiment and government support.

Economic Mobility and Public Opinion **Catherine Wilson, *American National Election Studies***

How does economic mobility relate to political attitudes and behavior? The American National Election Studies recently fielded a new set of questions about respondents' current, past, and anticipated future prosperity. We use these data to investigate two general research questions: How does past experience transitioning among low, middle, and upper incomes relate to political opinions? And how do expectations about one's chances of being poor, comfortable, or wealthy in the future relate to those opinions? We investigate these 'pocketbook' considerations of economic mobility as they relate to perceptions of the economy, blame attribution for poor economic performance, presidential approval, party identification, policy preferences, and presidential candidate preferences. We show how those who have experienced or who anticipate an upward trajectory to their financial well-being differ from those whose experiences or prospects have not been, or do not seem, as favorable, and we characterize the relative magnitude of these effects compared to other attitudinal and demographic variables.

Who Counts as White Working-Class? A Proposal for a New Approach **Daniel Cox, *Public Religion Research Institute*; Juhem Navarro-Rivera, *Public Religion Research Institute*; Robert P. Jones, *Public Religion Research Institute***

The influence of the white working-class on American culture and politics is difficult to overstate. Although arguably facing a decline in political clout, white working class Americans still retain an outsized influence in many important battleground states. Their support was pivotal for Obama in states like Michigan, Ohio and Pennsylvania. Yet despite this, there has been a glaring lack of consensus about the best approach to measuring this important group. In different works, the white working-class are defined in terms of income (Bartels 2008; McCarty, Poole, and Rosenthal 2006), occupation (Edsall 2007; Spitzer 2012), or some combination of these (McTague 2012; Teixeira and Abramowitz 2008). These different definitions often lead us to draw different conclusions about the political attitudes and behavior of the white working-class Americans. In this paper we compare several different definitions of the white working-class to a new definition developed from an original large (n=3,000) national survey of Americans. We define the white working-class using a combination of race, education, and an occupational proxy (people who are paid hourly or by the job). This definition is parsimonious and replicable and better captures the complex social, economic, and political realities of this oft-mentioned but often misunderstood group of Americans. This new approach provides a more complex picture of the working class in terms of their politics, economic outlook, and cultural traits.

The Employment Outlook of Low-Wage Workers in America

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

With a sluggish economy and shrinking middle class, there are many reasons to be concerned about the current status and future opportunities of lower-wage earners in America. The Associated Press-NORC Center for Public Affairs research, with funding from the Joyce Foundation, conducted a representative, multimode survey with 1,606 lower-wage workers in America to measure their opinions about their economic outlook, working conditions, and opportunities for advancement. The survey targeted employed individuals in jobs that pay less than \$35,000 per year. Findings from the survey reveal that lower-wage workers in American are struggling to get ahead, both inside and outside the workplace. Compared to the general population, more lower-wage workers feel that the country is headed in the wrong direction. Three-quarters of lower-wage workers report being worse off than they were four years ago and report worrying a great deal about many aspects of their personal financial situation. Inside the workplace, pessimism cuts across numerous aspects of their employment outlook with majorities seeing little opportunity for promotion or viewing that their current job will help them advance their long-term career goals. The data also reveal that this general and job related pessimism is especially high among white low-wage workers (even when controlling for other political, social, and demographic factors). In spite of a pessimistic outlook, lower-wage workers are generally satisfied with their jobs and working conditions, and a majority feel like their employer values them for the work they do. Findings do show that job training may be one solution to overcoming pessimism and feelings of being stuck in a dead-end job with a majority of workers who have participated in employer-sponsored job training programs and benefits reporting that training and education are important for moving ahead in their careers.

Seeing Red: The Politics of Regulations

Debbie Borie-Holtz, *Rutgers University*; Stuart Shapiro, *Rutgers University*; Michael Wong, *Rutgers University*

The role of regulation has been a central point of the recent presidential campaign and several gubernatorial contests. Regulations have been criticized as 'killing jobs' and hurting the economy while their defenders point to the benefits of a strong regulatory regime. Claims on either side of the debate are backed with limited evidence. While this is not unusual for claims in the political arena, academic examination of the effects of regulation has also been limited. In a unique dataset of environmental regulations, we examined whether regulatory burden hurts the economy or the business climate in five contiguous Midwestern states over the past decade. While the empirical data suggests NO, we then conducted a random survey of business leaders in these states to assess if regulatory criticism has any standing within the regulated community. At the outset, we conduct a list-experiment technique to measure whether regulations are considered a major problem among businesses, particularly given the attention paid to the policy issue by candidates and elected officials alike. While other national surveys suggest otherwise, we drill down further to see if regulatory burden is a 'real' or 'perceived' threat to businesses. If real, we attempt to determine if the threat is different among certain sectors of businesses, company size or gross revenues. If perceived, we look for reasons to explain in what ways regulations are considered harmful or economically threatening to business managers and owners.

Saturday, May 18
1:15 p.m. – 2:15 p.m.
Poster Session 3

1. Watch Your Language!: The Impact of the Survey Language on Bilingual Hispanics' Response Process

Meryem Ay, *University of Nebraska – Lincoln*; Wendy Gross, *GfK Knowledge Networks*; Curtis Cobb, *GfK Knowledge Networks*; Randall Thomas, *GfK Knowledge Networks*

Cross-cultural studies have been the focus of researchers creating and analyzing globally comparative data. However, the construct validity of the questions across cultures is a concern for data quality. Survey questions should have the same meaning and sentence structures across the languages. Current efforts are inclined to standardize survey questions across multiple languages, yet the impact of language itself as a potential confound remains untested. Linguists debate if language shapes individuals' thoughts and judgments or if a universal language of thinking exists (Whorf & Carrol, 1956; Chomsky, 1976). Given that increasing numbers of surveys are conducted in multiple languages, the debate among linguists introduces practical, yet untested, concerns. If the language that people use affects their thoughts and judgments, inter-language differences in multiple language surveys reflect not only true cross-cultural differences in attitudes and behaviors but also differences created by language itself. It is hypothesized that completing a survey in a specific language will prime respondents into thinking in a culture-specific way. Bilingual Hispanics are ideal subjects for this study because of their bridge between both non-Hispanic and Hispanic cultures and their ability to communicate in both English and Spanish. A Web-based survey experiment was conducted on a sample of 620 respondents from GfK's KnowledgePanel Latino to determine how the survey language would influence responses. Bilingual Hispanics were randomly assigned to receive questions in English (n=156) or Spanish (n=155) along with two control groups: English-only Hispanics (n=154) and Spanish-only Hispanics (n=155). Differences in acculturation between the two bilingual groups were examined to ensure randomization occurred properly. Even after controlling for demographic differences, preliminary analysis indicated differences between the two bilingual groups on topics related to self-efficacy. The differences are evidence of language priming and have potential implications for the data quality of multi-language, multi-cultural, and cross-national survey work.

2. Movers and Shakers: Discrepancies Between Cell Phone Area Codes and Respondent Area Code Locations in RDD Samples

Carol Pierannunzi, *Centers for Disease Control and Prevention*; Machell Town, *Centers for Disease Control and Prevention*; Lina Balluz, *Centers for Disease Control and Prevention*; William Garvin, *Centers for Disease Control and Prevention*; Mansour Fahimi, *Marketing Systems Group*; David Malarek, *Marketing Systems Group*; Ashley Hyon, *Marketing Systems Group*

Invariably, a portion of all cellular RDD sample telephone numbers reach individuals who reside outside of the area they are expected to reside—a discrepancy that only widens as the target geography becomes smaller down. In 2011, the Behavioral Risk Factor Surveillance System (BRFSS) found that on average about 8% of cellular calls reach individuals who reside outside of their sample states. The extent of this discrepancy ranged from a low of 4% in Mississippi to a high of 48% in the District of Columbia. By appending a variety of ancillary data about the location and demographic composition of rate centers

associated with each cellular number for the 2011 BRFSS, this research attempts to quantify and explain some of the observed discrepancies. Using multivariate analysis techniques, rate center characteristics are identified that may predict which numbers within a sample are more/less likely to reach respondents outside of their sample states.

3. Improving the Quality of Proxy Reports

Jennifer Edgar, U.S. Bureau of Labor Statistics

The Consumer Expenditure Quarterly Interview Survey (CEQ) asks one respondent to report expenditures made by an entire household. The CEQ has long identified this type of proxy reporting as a potentially significant source of underreporting. There are two likely reasons for these omissions: knowledge and recall. Lack of knowledge, stemming from the fact that participants may not know about all purchases by other household members, cannot be corrected through revisions to survey questions. The second reason, that participants may forget to consider other household members, may be able to be addressed through the survey design. A small lab study (n=20) explored the feasibility and effectiveness of collecting information about each household member at the beginning of the study, and using that information to add prompts in relevant sections of the survey. The study found this approach to be effective. All participants were able to provide information specific to other household members upfront, and after hearing the prompts reported an average of \$182 of additional expenditure reports; a 6 percent increase in overall reporting. This presentation will explain the method used and give an overview of the results.

4. Multi-Method Pretesting of Multilingual Survey Items

Cynthia Helba, Westat; Gina Shkodriani, Westat; Jasmine Folz, Westat; Martha Stapleton, Westat; Gordon Willis, National Cancer Institute

Cognitive interviewing and behavior coding are methods often used for pretesting survey questions prior to administration (e.g., Fowler & Cannell, 1996; DeMaio & Rothgeb, 1996; Willis, 2005). In single-language surveys, the two methods are frequently used together because they have different strengths (e.g., Census Bureau, 1995; Willis, DeMaio & Harris-Kojetin, 1999). A few studies also have reported on use of multiple methods to pretest translated questions (e.g., Napoles-Springer et al., 2006). Many of these studies use the number of problems identified through each method as a way to compare the testing methods (e.g., Presser and Blair, 1994); our study compares the specific types of problems identified by the two methods. This project used cognitive interviewing and behavior coding to pretest Chinese (Mandarin and Cantonese), Korean, and Vietnamese translations of the U.S. Tobacco Use Special Cessation Supplement to the Current Population Survey. We used 11 items translated into four languages. These 11 items were not revised between cognitive interviewing and behavior coding. Some had problems identified in cognitive interviewing and others did not. This project was an unusual opportunity to compare the types of problems identified using exactly the same items rather than items that had been revised between the cognitive interviewing and the behavior coding. Our analysis begins to assess whether multi-method pretesting can make a substantial contribution to multilingual survey development. We determine the types of problems identified at each stage of the pretesting using an established coding system for identification of questionnaire problems for the entire group of respondents. We then consider if the identified types of problems varied between language groups. Because the team also debriefed the behavior coders after the pretesting was completed, we will describe how these debriefings allowed the coders to act as “cultural interpreters” and further inform recommendations for revising the questions.

5. Targeted Data Collection Efforts for NASS's Quarterly Agricultural Survey Based on Nonresponse Classification Tree Models

Kathy Ott, *National Agricultural Statistics Service*; Melissa Mitchell, *National Agricultural Statistics Service*

In order to combat rising nonresponse rates, the National Agricultural Statistics Service (NASS) developed nonresponse propensity models to identify potential nonrespondents prior to data collection for the Quarterly Agricultural Survey. Classification tree models based on auxiliary data were used to rank order operations on their likelihood to be a nonrespondent. Earlier models were developed at the national level and now NASS has shifted focus to state level models. These scores can be used to enhance data collection efforts. Using the nonresponse prediction for each operation, targeted data collection methods were developed and tested to determine if specific methods directed at operations that had a low propensity to respond would increase response for the survey. The ability of the model to predict nonresponse propensity, as well as results from the targeted data collection methods, will be discussed.

6. Identifying and Addressing Response Inconsistency

Ashton Jacobe, *Fors Marsh Group*; Sarah Keaton, *Fors Marsh Group*; Luciano Viera, *Fors Marsh Group*

Measurement error occurs when a respondent's answer is inaccurate or imprecise. One common manifestation of measurement error is response inconsistency, where respondents provide survey responses that seem incompatible with or contradict their other responses. Response inconsistency is thought to occur when questionnaires are completed without full comprehension of the items. It is particularly problematic in self-administered surveys, due to limitations of implementing active follow-up and/or clarification strategies such as additional definitions, examples, or additional instructions. Survey research typically examines survey design features that contribute to response inconsistency, such as mode of administration, question type and wording, and interview setting. However, much less research has focused on identifying the contribution of respondents themselves to these errors. Despite efforts to develop instructions that communicate clear expectations and motivate high quality responses from all participants, these manipulations typically vary in their effectiveness in reducing response inconsistency across respondent types. This is consistent with past studies that have found that have found relationships between respondent demographic characteristics (e.g., age, gender, etc.) and the quality of survey responses. Additional research that takes a more holistic approach to response inconsistency is needed that examines both the survey design- and person-level causes as well as how they should be addressed. To this end, the present investigation uses multiple study examples to examine the: 1) Methods for identifying response inconsistency; 2) Survey design features that lead to greater response inconsistency; 3) Types of respondents that are more likely to respond inconsistently; and 4) Measures taken to reduce the impact of response inconsistency on measurement error. Results and implications for existing survey practice along with directions for future research will be discussed.

7. Controlling for Acquiescence in Comparative Cross-National Research: The Importance of Using Measurement Equivalent Country Clusters

Eva van Vlimmeren, *Tilburg University*; Guy Moors, *Tilburg University*

This paper addresses the situation in which an acquiescence response behavior has differential impact on cross-national differences in attitudes depending on the type of culture to which a national culture can be allocated. Acquiescence is a tendency among certain respondents to agree with question items irrespective of the content of the items and is generally recognized as a source that might bias cross-cultural comparisons. Frequently in

cross-cultural research there are problems regarding measurement invariance, i.e. the level of comparability of measurement models across cultures, we therefore grouped countries according to their homogeneity in measurement. Using data of the European Values Study of 2008, we will demonstrate that the correlation structure of a set of conceptually balanced items defines several clusters of countries that are internally homogeneous but can externally be quite diverse. Interestingly, the different clusters display a distinct reaction to controlling for ARS in the model. For instance, in the Western European cluster country differences in attitudes did not substantially change when controlling for acquiescence, whereas in the other clusters changes were pronounced. Our findings have important implications for comparative research in the sense that even a response style factor such as acquiescence can have a distinct meaning across cultures with various impacts on how it disturbs the measurement model. It also demonstrates that clustering countries according to their similarity in correlations within a given set of items might be a tool to identify measurement equivalent sets of countries in which comparative research is possible. Some practical guidelines as well as implications for further research are presented as well.

8. A Practical Approach for Identifying Engagement-Level Segments and Developing Differentiated Acquisition and Retention Strategies

Jack Fentress, *Data Recognition Corporation*; Herbert Baum, *Data Recognition Corporation*; Colleen Rasinowich, *Data Recognition Corporation*

Whether considering products, candidates or policies, individuals make choices. As researchers, a goal is to develop strategies that best impact individual choice and our offer's share of preference. Depending on the context, preference share may be translated as market share, votes or policy endorsement, but the analytics are similar. This presentation will provide an overview of our use of choice models to increase preference share. Standard analytic approaches tend to analyze populations in aggregate and identify universal drivers. Similar to the work of Fred Reichheld and the use of Net Promoter Score (NPS) in the commercial sector, we advocate the establishment of segments, differentiated by quantifiable levels of preference. The definition of these groups is flexible, but needs to segment respondents along a continuum of preference. Utilizing choice models and respondent-level analytics, we address two issues central to preference studies. The first is retention and how one can best strengthen preference among current supporters/customers. The second is acquisition or how one can best capture new supporters/customers. It's a balancing act between the retention of supporters and the need to revise that offer to acquire new supporters. The delineation of these segments and accounting for their unique requirements is recommended. We will present two analytic approaches that effectively address acquisition and retention strategies. Relative Strength of Preference (RSP) scorecards are effective for identifying acquisition strategies. Quantifying respondent-level gaps on key items results in the identification of those items that are most impactful for achieving overall preference. For the identification of retention strategies, Power/Penalty and Reward Quadrant Maps are highly effective. Utilizing logistic regressions, these maps identify which items have the greatest upside (reward), downside (penalty) or both (power). Our intent is to provide participants with several analytic approaches used in the commercial sector that have viable application for policy analysis.

9. Measuring Messy Concepts Without Creating Messy Questionnaires: The Case of Gender

Alian Kasabian, *University of Nebraska-Lincoln*

Researchers are often interested in the impact of gender on their variables of interest, yet use measures of sex category in their analyses. Sex and gender scholars are highly critical of this practice, due to the range of gender behaviors and experiences that are unrelated to

biological sex as labeled at birth and because there is growing visibility of people who do not identify as male or female. Yet for most surveyors, categorizing people as male or female is the most practical option because other gender measures tend to be very lengthy (as with psychological scales) or are better suited to qualitative work. To make real gains in incorporating gender into our understanding of the social world, researchers need a more nuanced and informative measure of gender that is not overly burdensome for respondents and does not require inordinate amounts of space in a questionnaire. In this paper, I present one such measure. In 2011, the Nebraska Annual Social Indicator Survey of residents aged 19 and older (n=906, AAPOR RR1=36.3%), provided respondents with a visual analog scale (VAS) labeled “completely feminine” on one end, and “completely masculine” on the other. Respondents were asked to place themselves, their spouse/partner (if applicable), and society’s ideal woman and ideal man on the scale. Thus, the scale provides an interval level measure of gender identity. Preliminary analyses indicate that respondents in the middle of the scale rate themselves significantly differently than their more feminine and masculine counterparts on a number of attitudinal measures (competence, political leanings, feminist identification, etc.), suggesting that the commonly used sex category measures are missing important variation. Additional analyses will assess the predictive validity of this gender measure. The paper will also discuss the difficulties of using a VAS for this construct.

10. Nonresponse Bias Analysis in a Cohort Study Incorporating Genetic Data **Daniel Loew, *Abt SRBI*; Mark Morgan, *Abt SRBI***

Post-Traumatic Stress Disorder (PTSD) is a mental health condition that afflicts many of the soldiers returning from service in Afghanistan and Iraq. Risk and resilience factors for PTSD are not well understood. Longitudinal research is being conducted to study the mental health trajectory of soldiers who have been deployed to combat situations and those who have not. It is critical to the interpretation of the results that study attrition is minimized and that bias over time is identified and adjusted for. The Ohio National Guard (ONG) cohort consists of ~3,000 members of the Ohio Army National Guard interviewed annually by telephone. Each member was also invited to submit a saliva sample for genetic analysis. The key questions that we will address in this methodological brief are: Are soldiers with more severe traumas more likely or less likely to continue participating? Are soldiers with less difficult service experiences more or less likely to continue? How do these potential biases affect our ability to identify the factors that prevent or promote the development of PTSD and other mental health problems? This methodological brief will examine the factors that are associated with attrition for the survey and participation decisions regarding the optional genetics study.

11. Four Experiments for the 2011 Diary of Consumer Payment Choice **Kevin M. Foster, *Federal Reserve Bank of Boston***

The Diary of Consumer Payment Choice (DCPC) is a new data product from the Federal Reserve Banks of Boston, Richmond and San Francisco. In 2010 and 2011, we conducted two pilot diaries, in which diarists reported all transactions--purchases and bill payments--and cash management activity over a three-day period. Respondents recorded their activity in a paper diary and then reported the results in a nightly online survey, which included additional questions. To prepare for the full implementation of the DCPC in 2012, we conducted four experiments concerning key survey methodology issues in the diary program: 1) Does using mixed modes affect the number of transactions reported? We asked some diarists to mail back their paper diary for an additional incentive. 2) Do new or experienced diarists report larger numbers of transactions? We feared that experienced diarists may suffer from diary fatigue or conditioned underreporting. 3) Do diarists who take the associated survey before their assigned diary period report different numbers of transactions than those who take the survey after their diary period? In the 2010 pilot study,

we insisted that all diarists take the survey first. 4) Does having extra 'lead time' affect the number of transactions and the amount of cash reported? Diarists receive their diary packet one, two, or three days ahead of their assigned diary start date based on the day of the week of the start date. The answer to each of these questions is 'No'. These results have the potential to save money (fewer incentives paid) and administrative effort (no need to remind diarists to take the survey first). In addition, the experimental outcomes show that we are not biasing our results by including both new and experienced diarists, nor by changing the lead time on receiving the diary packet.

12. Authorizing Health Record Linkage in Survey Research

Mindy Hu, *Mathematica Policy Research*; Ronghua (Cathy) Lu, *Mathematica Policy Research*; Anna Situ, *Mathematica Policy Research*

Linking administrative and survey data is becoming increasingly popular in health services research. Linking survey and medical claims data enables researchers to examine the interactions between disability, chronic disease, health care use, cost, and patient experiences with the health care system. Evidence suggests that participant characteristics—such as age, health status, and health care use—influence the likelihood to authorize data linkage; however, results are mixed regarding the most important variables and the direction of the effects (Beebe et al. 2011; Dunn et al. 2004; Harris et al. 2005; Huang et al. 2007; Knies et al. 2012). The enactment of the Health Insurance Portability and Accountability Act (HIPAA) of 1996 could help explain these mixed results. In the United States, the HIPAA Privacy Rule imposes requirements on obtaining authorization that could affect rates of authorization. Few population-based studies have examined the interplay of participant characteristics and authorization to link data in the context of HIPAA regulation. The 2012 Autoworker Health Care Survey is a self-administered mail survey of approximately 13,000 active and retired autoworkers and their spouses/partners consisting of 1) a health questionnaire; and 2) a request for written authorization (which meets HIPAA regulations), enabling researchers to link survey responses to medical claims data. Mathematica Policy Research conducted the survey for the National Institute for Health Care Reform. This paper will examine the influence of self-reported health, health care use, and demographic characteristics on rates of authorization to link survey data to medical claims data. We will use logistic regression to examine associations between individual characteristics and authorization outcome. We will also examine potential bias due to differences in authorizers and non-authorizers and discuss the resulting implications for survey design.

13. Can a Verbal Prompt About Importance 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 non-response 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 experimentally tests whether a verbal prompt about the importance of answering the demographic questions improves item response rates (relative to the version without the prompt) 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. 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 a verbal prompt about importance is a viable strategy for reducing item non-response of demographic items.

14. An Experiment to Improve Spanish Language Response Rates to a Mail Questionnaire

Andrew Caporaso, *Westat*; David Cantor, *Westat*; Aaron Maitland, *Westat*; Bradford Hesse, *National Cancer Institute*

The Health Information National Trends Survey (HINTS) is a national health communication mail survey sponsored by the National Cancer Institute (NCI). In the first cycle of HINTS 4 non-responding households were mailed both an English and Spanish questionnaire in the second mailing if their address was linked to a Hispanic surname and/or was in a linguistically isolated (LI) area as indicated on the frame. This strategy yielded a sample which was 8.5% Hispanic, which was significantly lower than ACS figures. Compared to prior telephone versions of HINTS, significantly fewer surveys were completed in Spanish. Since cycle 1, Brick et al. (2012) have reported on a different mailing procedure that was tested with a short screening survey on education. This test found significantly more returns of Spanish language surveys, as well as more Hispanic respondents, when compared to the cycle 1 HINTS procedure. The purpose of this paper is to test whether these results generalize to HINTS, which is a long survey (about 20 pages) on a topic that is less salient than that tested by Brick et al. The paper will report on the results of an experiment that was carried out in cycle 2 of HINTS 4 which compared two different mailing methods intended to reach more Spanish speakers and Hispanics. In the first condition, based on Brick et al., about 2,000 respondents were sent both a Spanish and English questionnaire in all mailings. In the second condition, about 10,000 were sent a Spanish and English questionnaire in all mailings only if the household was linked to a Hispanic surname and/or LI area. The presentation will report on the results of the experiment with respect to the number of Spanish language returns, the percentage of respondents identifying as Hispanic and overall response rates.

15. All in the Family? Who Do Respondents Include When Responding to Telephone Status Items

Josiane Bechara, *NORC at the University of Chicago*; Vincent Welch, *NORC at the University of Chicago*

The benchmark study for telephone status in U.S. households is the National Health Interview Survey (NHIS) published by the National Center for Health Statistics. The NHIS is an area probability survey where data are collected face-to-face in an interview that lasts for nearly an hour. Telephone status on this survey (i.e., wireless-only, wireless-mostly, landline-only) is established through responses to a survey item that asks 'Of all the telephone calls that you or your family receives are...' In the context of the NHIS interview, researchers believe that respondents clearly understand what the term 'family' should include (See Blumberg and Luke, 2012). This item has been employed in a number of studies that are conducted over the telephone. It is not clear that respondents in the telephone setting understand the term 'family' in the same way that NHIS respondents do. The current research explores telephone respondents understanding of the term 'family' in this telephone status option. We employed in-depth probing in a cognitive interview setting in order to understand the level of agreement between respondents' household rosters and the set of individuals whom they included in their 'family' when responding to this item. We

found that respondents made errors of inclusion and exclusion in their 'family' composition. Replacing the word 'family' with the 'household' dramatically reduced the number of errors and led to increased reliability. Further probing revealed that respondents' self-generated definition of 'household' was also in line with Blumberg and Luke's (2012) intended meaning. Implications for future dual-frame RDD studies are discussed.

16. The Expansion of Survey Research into Educational Strategy Consulting: An Example of How Universities Can Increase Retention Rates With the Use of Surveys and Personality Tests

Thomas Lamatsch, *Monmouth University*; Tyler Breder, *Monmouth University*; Andrew Bell, *Monmouth University*

In order for survey research to have a sustainable future it is important to branch out and cooperate more closely with other fields and break into new areas. While survey research organization have done large scale studies of education systems for decades they are mostly absent in the field of education consulting which is dominated by MBAs and researchers with education degrees although we should play a more serious role. One of the major problems universities struggle with today is retention and survey research could assist in that issue similar to the way Gallup assists their clients in picking employees who are the right fit for their companies. This paper will, however, turn the premise around and not look for the right student to fit the university but the right fit in terms of approaches to teaching geared for their students. Universities have long acknowledged that few students leave because they struggle academically; instead they leave because it is not what they expected. This study will test if schools could increase retention rates by offering more flexible programs and closer advising to students based on students' characters and temperaments. Companies worldwide use the Meyer-Briggs typology test (MBTI) to create the ideal atmosphere for their employees to succeed. This study will conduct a survey of 500 randomly selected students who will answer questions modeled after the MBTI as well as question about how happy they are with their choice of college as well as their preferred form of "education delivery," i.e. lectures, seminars, independent studies, online classes etc. The results can then be used to create simple tests that advisers should use to not only advise their students academically but advise them which type of class they will most likely succeed in.

17. Immigration à la GCC: Support and Opposition to the Kafala System in Qatar

Abdoulaye Diop, *Social and Economic Survey Research Institute, Qatar University*; Trevor Johnston, *University of Michigan*; Kien T. Le, *Social and Economic Survey Research Institute, Qatar University*; John L. Holmes, *Social and Economic Survey Research Institute, Qatar University*

Since the 1950s, immigration in the Gulf Cooperation Council (GCC) countries has been uniquely governed by the Kafala or sponsorship system. The Kafala provides the legal basis for the residency and employment of migrant white-collar and blue-collar workers in these countries. Today, despite growing criticism from human rights organizations, little effort has been made to ameliorate the difficult working and employment conditions of these migrant workers in the GCC countries. While the existing literature is abundant at the country-level, combining macro analysis and ethnographic narratives to describe the abuses and human costs, we know little about public opinion towards the Kafala system. Capturing this public opinion is critical to understanding the GCC countries' failure to enact vital reforms. In this paper, we study this issue using data from two nationally representative surveys in Qatar. We begin by exploring the native Qataris' attitudes towards migrant workers in general and the determinants of support for or opposition to reform. Drawing on a survey experiment, we then exploit a matching design to evaluate the effects of priming and prejudice on the

Kafala's reform. Finally, we draw some conclusions about the results with respect to the future outlook of the region.

18. Evaluations on a New Methodology of the Turkish Consumer Survey

Türknur Hamsici Brand, *Central Bank of Turkey*; Ece Oral, *Central Bank of Turkey*

This study investigates the methodology of a redesigned monthly Turkish consumer confidence survey conducted by Turkish Statistical Institute and the Central Bank of Turkey in an effort to calculate the consumer confidence index for Turkey. Since the start in December 2003, the data collection method of the Turkish Consumer Survey has been a face-to-face survey annexed to a Labor Force Survey panel design. In addition to data collection method, the redesigned survey has implemented a different sampling method. Recently, the updated survey has been at the pilot period for twelve months. Comparisons between old and new survey at a design-based perspective are made. Reasons for possible measurement errors and biases are evaluated. Consumer surveys are usually aimed at including questions to form consumer confidence indices, and maintain comparisons across countries via standardized questionnaires and indices. The redesigned Turkish survey meets the requirements of the European Commission quality dimensions for a future approval. Consumer confidence indices are used as economic indicators for forecasting household consumption expenditure, consumer behavior in general, and the country's economic situation. The data are also used in political decision-making processes. In this regard, consumer confidence surveys are useful and widely accepted tools for gathering information about common people's expectations over time (Ludvigson, 2004). Some indicators derived from Turkish consumer confidence survey are used in macroeconomic analysis and forecasting. Given the significance, the choice of survey methodology used is central to maintain efficient results for economic and political environment. The redesigned survey is expected to be a good asset to improve the macroeconomic indicator characteristics of Turkish consumer confidence index.

19. In Search of More Granular Likely-Voter Models for Low-Turnout Elections: The Case of the 2013 Florida and Ohio Primary Elections

Clifford Young, *Ipsos Public Affairs*; Neale El-Dash, *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. In the specific case of the 2012 Ohio and Florida primaries, we confronted these issues. First, 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 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 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 were able to discriminate voters in one-percent intervals from 0 to 100%. Second, by employing political variables, our model captured 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 13,000 interviews collected for the Reuters-Ipsos 2012 primary

polls in OH and FL. To measure performance, we will employ the Average Absolute Difference between the survey estimates and election results.

20. The Effectiveness of Follow-Up Interviews in Reducing Item Nonresponse Bias in Mail Surveys

Sandra L. Clark, U.S. Census Bureau; Deborah H. Griffin, U.S. Census Bureau

Research has demonstrated that survey managers need to consider factors other than response rates when assessing survey quality. When considering nonresponse, quality is a consequence of the adjustments that a survey makes and the similarity of survey nonrespondents and respondents, more than the level of nonresponse. While much of the research in this area has focused on unit or survey nonresponse, item nonresponse involves parallel concepts and concerns. We generally assume that a low level of item imputation is a good predictor of the quality of survey estimates. This paper assesses if efforts to reduce levels of item nonresponse in the American Community Survey (ACS) are successful in reducing nonresponse bias. The ACS achieves high levels of item response because data collection includes special efforts to follow up on incomplete responses. Evaluations have demonstrated the effectiveness of this follow up effort in reducing the national-level item imputation rates. These evaluations have not assessed the reduction in nonresponse bias that the ACS achieves by converting a subset of item-level nonresponses to responses. Recent analysis of this follow up operation provides us with important information about our ability to obtain responses for items that respondents left blank on ACS mail-returned questionnaires. Using data from the 2010 American Community Survey, this research identifies the specific items that follow up efforts are successful in converting and those that once left blank stay blank. In addition, to assess nonresponse bias reduction, this paper compares the values of the originally missing, converted responses to the values reported without follow up. By closely examining the ACS's mail return follow-up operation, this project will broaden our knowledge of item nonresponse bias in mail surveys and help us define the items that benefit most from follow up efforts.

21. Conducting "Issues" Surveys Using Automated (IVR) Polls: The Case of the National Leadership Index

Seth A. Rosenthal, DataDoc Research Consultants; Owen Andrews, Center for Public Leadership, Harvard Kennedy School

The use of automated (IVR) polling methods to conduct issues-based surveys is controversial. Issues-based survey questions are often more complex than the candidate-choice questions typical of IVR polls. Some critics suggest that only live-caller interviews can provide valid assessments of public opinion on complex issues. We evaluated the validity and effectiveness of IVR-based issues polling using data from the National Leadership Index (NLI). The NLI is an annual survey in the U.S. of public opinion toward the nation's leaders. It is conducted by the Center for Public Leadership at the Harvard Kennedy School in collaboration with Merriman River Group. It assesses and indexes opinions about national leadership across 13 key sectors of public life. From 2005-2010, the NLI was conducted as a live-caller survey. In 2010, we tested a pilot IVR version of the NLI, which allowed for direct comparison of the two methods. Since 2011, the NLI has been conducted as an IVR survey. Overall, our data indicate a nearly seamless transition from live-caller to IVR methods. Two areas, however, merit closer examination. First, there was an increase in endorsement of extreme responses in the IVR version, particularly on the most divisive questions. However, the mean responses for these questions were not affected. This may indicate that the increase in extreme responses accurately reflected respondent opinion after the moderating effect of a live interviewer was removed. Second, the percent of "not sure" responses increased marginally throughout the survey. This was likely due to the

inclusion of an explicit “not sure” option for each question, necessitated by the IVR methodology. However, some argue that including “not sure” anchors generally increases the external validity of public opinion surveys. Overall, results for the IVR version of National Leadership Index suggest that IVR can compare favorably with live-caller methods for conducting issues-based surveys.

22. Is Interactive Voice Recognition a Viable Mode of Data Collection?

Adam Gluck, *Arbitron*

Arbitron uses a panel-based methodology to collect radio listening data, and produce media ratings in various markets around the country. The method for collecting this data is the Portable People Meter, a cell phone sized device that passively measures exposure to encoded audio in media. As each individual meter is carried by a unique panelist, we can associate the media that the PPM detects to the panelist who is wearing it, thus creating an electronic log of their listening. From that we can estimate who was listening to radio. After panelists leave a panel, we occasionally re-contact them to gather additional information via surveys. During the fourth quarter of 2012 and the first quarter of 2013, Arbitron will conduct one such brief survey. Households will first be surveyed via phone, with a brief survey administered automatically via Interactive Voice Recognition (IVR). Households will be sent one follow-up IVR call as well, and they may call back a special number to administer the IVR survey at their leisure. The survey will consist of two questions. In this paper, we will seek answers to the following questions: 1. What type of response rate does an IVR survey yield? 2. What are the characteristics of responding vs. non-responding households, with regard to phone type (cell vs. landline), income level, size, and presence of children? Additionally, we will also present information about the legal and logistical challenges of administering an IVR survey.

23. The Effectiveness of Forgiving Introductions and Response Options for Reducing Social Desirability Biases in Reports of Health-Related Behaviors

Hanyu Sun, *Joint Program in Survey Methodology*; Rebecca Medway, *American Institutes for Research*

As the obesity epidemic continues to rage, it is becoming increasingly important to collect accurate information about people’s health-related behaviors. Unfortunately, it can be difficult to get survey respondents to provide truthful responses about these topics. One method researchers have proposed as a way to reduce such social desirability biases is adding a forgiving introduction to the question stem. It is hypothesized that forgiving introductions reduce both the intrusiveness of the question and respondents’ concerns about the negative consequences of giving a truthful response. However, the few experimental studies that have tested their effectiveness have produced mixed results. One explanation for these mixed results is that many studies utilize vague introductions that respondents do not find very convincing (e.g., “Some people want to exercise, but they just can’t find the time”). We hypothesized that offering concrete, scientific statements would be a more effective approach (e.g., “A recent study conducted by the Center for Disease Control indicates that almost one-third of adults do not exercise on a regular basis”). Additionally, the previous studies rarely experimentally manipulated both forgiving introductions and forgiving response options simultaneously. Finally, most existing studies have focused on reports of voting history and sexual behavior; the effectiveness of forgiving introductions and response options on reports of other health-related behaviors has not yet been investigated. To better determine whether, and when, forgiving introductions and response options are effective, we included a 5-item 3x3 question wording experiment in a national probability-based Web survey. The experiment varied both the authoritativeness of the forgiving introduction (authoritative scientific introduction vs. vague non-scientific

introduction vs. no introduction) and the use of forgiving response options (forgiving response options first vs. forgiving options last vs. no forgiving options). This presentation presents the results of this experiment.

24. Reaching Respondents Using an Address-Based Frame: Does a Nonreturned Mail Questionnaire Really Mean “No”?

Marla D. Cralley, *Arbitron*

Over the past ten years researchers have witnessed decreasing coverage and efficiency in traditional Landline RDD samples. To address this, Arbitron conducted experiments and began using cell-only and cell-mostly samples to supplement the traditional RDD samples. Finally during 2011, Arbitron moved to a total Address-Based sample frame in the 47 top media metros currently measured by Arbitron’s PPM service. The Arbitron PPM service passively collects Radio and Television media usage among an on-going panel of respondents. This system replaced the traditional paper Radio and Television self-report diaries previously used in these markets. Address-based sampling requires researchers to employ differing modes to contact potential respondent households effectively and economically. Arbitron uses an initial phone contact to reach resident addresses where Arbitron’s sample vendor is able to match to a phone number using secondary databases. Selected addresses where a phone match is unavailable are initially contacted using a screener questionnaire. This questionnaire is designed to confirm the address reached and collect demographic information and a telephone number. Selected households returning usable questionnaires are then contacted using the provided phone number for panel recruitment. Attempts to recruit a sub-sample of households that do not return usable questionnaires are made in person by Arbitron field representatives. This paper compares panel recruitment agree rates for households returning the initial mail questionnaire to those who did not return the screener. Recruited households will also be compared based on household demographics and quality of panel participation. This analysis will evaluate the benefit of making additional efforts to contact households not returning mail questionnaires.

25. Motivated Conservationism: Contingent Effects of “One Health” Framing on Conservation Behavior

Sungjong Roh, *Cornell University*; Katherine A. McComas, *Cornell University*; Dan Decker, *Cornell University*; Rickard Laura, *SUNY-ESF*

Recent years have seen growing attention to communicating about the interconnectedness of human, environmental, and animal health. Our research on “One Health” messages examines how framing wildlife diseases as not only resulting from wildlife behavior but also due to human and environment factors might influence conservation behaviors seeking to protect the natural environment (see Karesh & Cook, 2005 for a review). Yet, recent work in framing effects suggest potential boomerang effects (Chong & Druckman, 2007) when individuals who receive information opposing their belief system may not simply resist challenges to their views but instead strengthen their original, opposing position (e.g., Gollust, Lantz, & Ubel, 2009; Peffley & Hurwitz, 2007). Building on research (Kahan, Jenkins-Smith, & Braman, 2011) into the boomerang effects of message framing caused by individuals’ ideology-protective cognition (i.e., cultural cognition; DiMaggio, 1997; Douglas & Wildavsky, 1982), we investigate how message effects of a One Health frame and its counter-frame (i.e., blame wildlife behavior only) vary by citizens’ cultural values (Hierarchical-Individualists vs. Hierarchical-Communitarians vs. Egalitarian-Individualists vs. Egalitarian-Communitarians). We report on a Web experiment of N = 550 Americans who reported intentions to engage in conservation behaviors. Results varied markedly by frames and individuals’ cultural cognitions. Specifically, among Egalitarian Individualists, the One Health frame showed a boomerang effect: it reduced intentions to engage in conservation

behaviors compared to a control group, which did not read a message; however, the counter frame, which blamed wildlife behavior, led Hierarchical Communitarians to express greater intentions to engage in conservation behaviors compared to the control group. Our discussion focuses on theoretical and practical implications of the efficacy of One Health framing in messages seeking to increase conservation behaviors among a diverse public audience.

26. Vacant Housing Units and Other Out-of-Scopes Identified Across Data Collection Years of the General Social Survey (GSS)

Jodie A. Daquilanea, *NORC at the University of Chicago*; Katherine Dekker, *NORC at the University of Chicago*; Lauren Doerr, *NORC at the University of Chicago*; Ned English, *NORC at the University of Chicago*

The General Social Survey (GSS) provides a suitable environment in which to explore trends in vacancy and housing unit eligibility rates, as it has been conducted as a nationally-representative household sample over the past decades. The GSS, sponsored primarily by the National Science Foundation, biennially collects cross-sectional and panel data on the attitudes, experiences, and demographic characteristics of residents throughout the United States. The cross-sectional sample uses an address frame based on the United States Postal Service Delivery Sequence File (DSF), as enhanced through supplemental listing conducted by NORC staff prior to the start of data collection. NORC updates its national sampling frame for the cross-section component of the GSS every ten years in rural areas based on the decennial Census. Field interviewers visit cross-sectional housing units during data collection, and so determine their eligibility. Sampled housing units may then be identified as being vacant or not housing units and therefore out of scope. The 2004 and 2012 rounds of the GSS used newly-updated sampling frames, based on the newest released Census data. For this paper we will track trends in vacancy rates and housing unit eligibility rates across multiple years going back to 2000. Observed vacancy rates may increase as the sample frame ages. Further, vacancy rates in later years may have been affected by the 2008-09 economic recession; the GSS provides an environment in which we can observe these trends over two-year intervals. For rural areas that required in-person listing, we will also compare vacancy rates reported by the Census with vacancy rates calculated through GSS fielding. These findings will add to the body of knowledge about the effect of recency of updates in a study's sample frame upon vacancy rates calculated in its subsequent fielding.

27. Comparisons of Online Recruitment Strategies: Craigslist, Facebook, Google Ads and Amazon's Mechanical Turk

Christopher Antoun, *University of Michigan*; Chan Zhang, *University of Michigan*; Frederick G. Conrad, *University of Michigan*; Michael F. Schober, *The New School for Social Research*

Methods such as posting flyers in public places, placing print ads in newspapers and magazines, and posting online classified ads on Craigslist have been widely used to recruit research subjects. Recently, the rise of social media Websites (e.g., Facebook) and online services such as Google Ads and Amazon's Mechanical Turk (MTurk) offer new opportunities for researchers to recruit study participants. Although researchers have started to use these emerging methods, little is known about how they perform in terms of cost efficiency and, more importantly, the type of people that they ultimately recruit. Here, we report findings about the performance of four online sources for recruiting participants, in our case, iPhone users: Craigslist, Facebook, Google Ads and MTurk. First, we compare the cost and participant demographics associated with different recruiting sources. Next, we evaluate whether people recruited from different sources behaved differently in our screener

survey (a brief online questionnaire to collect participants' demographic information and to verify they are actually iPhone users). The findings reveal very different performance between two types of online recruitment strategies: those that "pull-in" online users actively looking for paid work (e.g., MTurk workers and Craigslist users) and those that "push-out" a recruiting ad to online users engaged in other, unrelated online activities (e.g., Google ads and Facebook). We find that (1) the pull-in recruiting strategy was more cost efficient (more respondents per dollar) than the push-out approach; (2) participants from the two pull-in sites (Craigslist and MTurk) were predominantly young presumably because the users are relatively young; (3) the two push-out recruiting sources, in contrast, seemed to have reached a more diverse user base. In addition, the pull-in strategy brought in participants who seemed more committed to the task and more willing to disclose personal information in the interview, than respondents attracted through push techniques.

28. Continuous Survey Improvement: Modeling Nonresponse in Real-Time to Optimize Sampling and Contact Procedures

Andrew Therriault, *Lightbox Analytics*

Disposition data is regularly used for post-survey adjustment, most commonly to reweight for representativeness, but a more proactive approach offers a chance to address these issues in real-time. We present an original method---'continuous survey improvement'---for using disposition data from surveys still in the field. Our technique is based on modeling nonresponse to initial survey attempts as a product of the various data available, including completed surveys, call metadata, and characteristics of the target population. Through the use of Random Forests, Lasso models, and other data mining tools, we can not only pinpoint which segments of the population are being missed, but also identify how best to correct the problem with changes in sampling or contact procedures. By addressing problems during the survey rather than afterward, the ultimate goal is to reach truly-representative set of respondents, rather than settling for weighted approximations. While our method is most obviously applicable to long-term or repeated surveys, (e.g., tracking polls, unemployment surveys), the same process could be applied in the course of one-off surveys as well.

29. The Effect of Stamped Return Envelopes on Re-Mailing to Non-Respondents

Scott A. McInerney, *Center for Survey Research*

Although it is more economical for researchers to use business reply return envelopes when sending out mail questionnaires, evidence has shown that stamped return envelopes improve response rates by several percentage points. This has been shown for initial survey mailings by Dillman and others, however, to date there is no published research addressing the effect of stamped return envelopes on response rates for second round mailings to non-respondents. Our experiment was designed to see if the benefit would persist in the second round. As part of the Indiana/Texas Tobacco Study at the Center for Survey Research at University of Massachusetts Boston, paper questionnaires were used to reach address based sample (ABS) without any listed phone number. After an initial mailing to 4,000 sample members, each including a \$1.00 incentive and a stamped return envelope, followed by a reminder postcard, we still had 2,630 non-respondents. For the re-mail of the survey instrument, we randomly assigned half the non-respondents to a stamped return envelope condition, and half to a business reply envelope condition. No incentive was included in the re-mailing. Comparing both groups, results show no significant difference between the rates of return (9.8% vs. 9.3%). As previous research indicates, the idea of stamped return envelopes may boost response rate for the initial mailing; however, it does not seem to improve the rate of return for the re-mail of potentially more resistant non-respondents. This research was funded by the National Cancer Institute, Grant #5R01CA151384.

30. Polling Post-Superstorm Sandy: Understanding the Social and Political Aftermath of the Hurricane in New Jersey

David Redlawsk, *Rutgers University*; Ashley Koning, *Rutgers University*; Elizabeth Kantor, *Rutgers University*; Caitlin Sullivan, *Rutgers University*

The entire Northeast and especially New Jersey suffered severe damage and loss from Superstorm Sandy in October 2012. Rendering many regions powerless and devastated and hitting soon before the election, the storm had serious social and political consequences for countless citizens – as well as implications for polling and the field of public opinion in the last days of presidential campaigning. In the storm’s aftermath a few weeks later as New Jersey slowly began to return to a “new normal,” the Rutgers-Eagleton Poll carefully captured citizens’ opinions in a Sandy-focused post-election survey on how the storm affected them both personally and politically. In terms of personal ramifications, this analysis looks at whether New Jerseyans were affected by Superstorm Sandy, forced to evacuate, sustained property and/or other damage, and suffered power outages. It also assesses interaction with and opinions on FEMA, the Red Cross, and citizens’ electric companies, as well as the state’s overall level of preparation. Politically, we investigate how Sandy impacted New Jersey voters on Election Day, whether it swayed their vote, how they viewed Governor Chris Christie’s and other political figures’ handling of the crisis, and what they thought of the highly publicized bipartisan visit between the governor and President Obama after the storm. This analysis provides a look into New Jersey opinions soon after Sandy’s aftermath by standard demographics such as income, race, and region, as well as by cell/landline telephone contact and day of interview.

31. Barking up the Right Tree: Surveys to Target and Analyze Animal Health

Danna L. Moore, *Social and Economic Sciences Research Center, Washington State*; Thom Allen, *Social and Economic Sciences Research Center, Washington State*; Rose Krebill-Prather, *Social and Economic Sciences Research Center, Washington State*

A significant issue for many animal health researchers is to define and get information from the very specific subgroup of the human population closely associated with an animal population that is at risk or that has a higher risk of injury, illness, nutritional requirements, and/or special performance requirements. This research discusses aspects of locating and finding a hard to reach group, owners of agility dogs and defining measurements of nutrition and health, incidence of injury and illness, and animal health practices. The incidence of one specific feeding practice is evaluated that is closely connected with infectious disease transmission between dogs and humans. This study examines the problem of a population within a population. A targeted large convenience sample, a general population survey, and social network recruitment are used to study incidence and to comparatively study this problem. New social media are used as an optional innovative framework for sampling, targeting, and evaluating complex health problems where the contactable population holds key information related to a subpopulation of interest.

32. Combining Local and National Cross-Survey Data to Estimate the Prevalence and Characteristics of Low Incidence Religious Groups in the New York Metropolitan Area

Daniel Parmer, *Cohen Center for Modern Jewish Studies*

One of the defining characteristics of the United States is its religious diversity and the traditions of civic involvement and service of many of the religious communities. However, the separation of church and state precludes the U.S. government from collecting data on the religious identification of citizens. An important source of estimates of the religious composition of the U.S. is surveys, such as the American Religious Identification Survey as

well as surveys commissioned by specific religious denominations. Single surveys as sources of estimation are problematic. Many include too few respondents to be able to describe reliably the low-incidence religious groups (those ranging from 1% to 10% of the population). Moreover, any individual survey contains systematic errors that arise from questionnaire construction, sampling, sponsorship, and “house” effects. This study seeks to overcome these challenges through the development of cross-survey analytic techniques that are similar in approach to standard meta-analyses. We have compiled data across more than 50 independent surveys of the New York metropolitan adult household population. Each survey was designed to provide a representative sample and each contained questions about current religious affiliation. Multilevel and advanced Bayesian techniques were employed to account for within survey clustering and to develop estimates of smaller groups, such as Jewish, Mormon and Muslim, as well as larger groups such as Catholic. Estimates were post-stratified across surveys on basic demographics such as age, sex, race and educational attainment. In addition, adjustments were made for the over- or under-representation of metropolitan areas across the sample of surveys. The results from this analysis expand on prior research by combining national and local data sources to estimate the prevalence and characteristics of low incidence religious groups at the metropolitan level.

33. Commemoration Matters: The Anniversaries of 9/11 and Woodstock
Amy Corning, *University of Michigan*

We investigate the effect of anniversary commemorations of September 11 and Woodstock on the American public’s collective memory or collective knowledge of each event. We are able to examine both the eighth and the tenth anniversary commemorations of the September 11 attacks (in 2009 and 2011), as well as the fortieth anniversary of the 1969 Woodstock Festival (in 2009). In an initial step, we used media analysis to identify the timing of commemorative activity surrounding the anniversaries. Our second step was to draw on data from surveys whose fieldwork dates corresponded to the anniversary periods, in order to compare respondents’ memory and knowledge of the events before, during, and after the commemorations. Our evidence shows that the percentage of Americans who consider 9/11 an “especially important” event is related to commemorative activity, and we likewise find that greater knowledge about the Woodstock festival is associated with commemoration of that event. In addition, the impact of commemoration on knowledge of Woodstock was greatest among those with lower levels of education. For memory of 9/11, we found that commemoration’s effects were stronger for blacks than for whites, suggesting that commemoration may enhance the salience of national, as opposed to racial, identity. These findings offer insights into the educative and evocative roles of commemoration.

34. The Prevalence and Impact of Self-Selection Bias and Panel Conditioning on Smoker Studies Using Established Internet Panels
**J.M. Dennis, *GfK Knowledge Networks*; Curtiss Cobb, *GfK Knowledge Networks*;
Michael Lawrence, *GfK Knowledge Networks*; Jordon Peugh, *GfK Knowledge Networks***

Given their many advantages (see Couper 2008; Fricker 2002; Chang & Krosnick 2009), it is not surprising that there has been increasing use of established Internet panels for household and individual level data collection. Internet panels are, however, susceptible to two potential drawbacks: self-selection bias and panel conditioning effects. Self-selection bias is a form of non-response and can occur if panelists non-randomly fail to participate in assigned studies or fail to answer specific questions within a study. Panel conditioning can occur if panelists’ responses in a study are influenced by participation in prior studies, such that panelists’ answers differ systematically from those of individuals not on the panel. Self-

selection bias and panel conditioning effects may be particularly likely to occur for individuals asked to complete many surveys on the same topic while a part of the Internet panel, such as what occurs with smokers and public health smoking studies. This study investigates the prevalence and impact of these biases on three smoking-related public health studies conducted using GfK's KnowledgePanel®, a probability-based Internet panel representative of the U.S. general population. Outcomes examined include measures of knowledge, behavior and attitudes and are estimated from selection models to disentangle conditioning from non-response. Initial findings suggest that while many questions related to attitudes, behaviors and knowledge are repeated across most smoking-related studies, exposure to prior smoking surveys was weakly correlated to respondent answers in two out of the three smoking studies examined. For example, panel conditioning effects were estimated to increase the prevalence of having ever tried to quit smoking from 62% to 63.2% (+1.2 points). Not surprisingly, willingness to participate in early studies, regardless of topic, is related to the likelihood of completing another smoking study. These results are reassuring that panel participation minimally impacts the reported attitudes and behaviors of respondents.

35. Voter Identification: Towards A Statistical Likely Voter Model

Jonathan Robison, *Greenberg Quinlan Rosner Research*; Masahiko Aida, *Greenberg Quinlan Rosner Research*

There has been much controversy in political punditry on the criterion for assessing whether a respondent will be a likely voter in an election. As is commonly known, the likely voter models many political public opinion researchers use are not statistical in nature, rather they are decision rules meant to define a universe that, apriori, researchers believe will constitute the electorate. Recent scholarship has made substantive critiques of likely voter models that use variables such as enthusiasm and political knowledge, and proposed differing methods for resolving biases that likely voter screens introduce. With declining response rates to surveys, developing an empirically rigorous and statistically grounded likely voter model will go a long way towards improving accuracy and limit bias in results. Today, pollsters using likely voter models rarely go back to validate its effectiveness, relying on gut instinct rather than hard data. Because the existing literature in this area is relatively sparse, uses older and less extensive data, as well as less rigorous predictive methods, we believe we can make a both scholarly and practical contribution to this area of research. Using sophisticated predictive modeling techniques, we intend to create a weighted algorithm to assess the likelihood a registered voter will vote, using data from a national survey to create a statistical decision rule that will provide researchers with a dynamic, rather than an ad hoc method to create a likely voter universe. Additionally, the novel dataset the authors assembled for analyzing likely voter screens includes data of the evaluations by calling house professionals instructed to rate the likelihood a respondent will turn out to vote on Election Day. Utilizing this novel survey question and survey micro-data, we plan to find an optimal likely voter screen.

36. Analyses of a Frame Based Telephone Survey in Mainland China

Shishi Chen, *The University of Hong Kong*

Fixed lines and mobile phones have been widely used as national telephone survey tools and there are many studies of fixed line and mobile phone survey methodology and comparing telephone surveys with other survey modes. This paper builds upon a great opportunity for methodological work on fixed line and mobile phone surveys in Mainland China, using a follow-up survey interviewing the respondents from a prior face-to-face survey. This is innovative. Understanding the challenges in fixed line and mobile phone surveys in Mainland China is a very topical issue in the field of survey research and the

results can be used to study survey errors and contribute to that literature as well as to improve the quality of survey fieldwork procedures. A database with telephone contact information for 4041 individuals was obtained from a household survey in Mainland China, for which the Social Sciences Research Centre of the University of Hong Kong was commissioned to conduct a follow-up telephone survey of the same individuals. The households were sampled randomly for the first wave national face-to-face survey and the individuals are respondents who left their telephone numbers after the face-to-face survey and accepted in principle a follow-up interview within two weeks. This paper analyzes the quality of the face-to-face database and the outcomes of the follow-up telephone survey. As the demographics of respondents and non-respondents were known from the database, studies of the influence of day, time, household demographics and individual demographics on the first and second contact attempt outcomes are undertaken using logistic regression. The findings include an effective calling design to improve telephone survey field work strategy and contribute valuable information for further studies in Mainland China. The impact of the interviewers' language skills on survey cooperation rate is also discussed.

37. Debating Tweets: An Analysis of Policy Choices on Twitter During the Dutch Pre-Election Debates

Bengü Hosch-Dayican, *University of Twente*; Kees Aarts, *University of Twente*

To what extent can social media be a relevant data source for the study of political representation? The present paper aims at providing some building blocks for answering this question, using data collected on Twitter during a 2012 election campaign. A commonly used measure of the quality of political representation is the congruence between policy preferences of the electorate and their representatives. Recent research has demonstrated, however, that measures of ideological and issue proximity between voters and parties based on survey data and content analyses of party programmes lead to contradictory findings on party representativeness (Thomassen 2012). This suggests that traditional methods of analyzing issue congruence should be accompanied by more comprehensive data. We aim therefore in this paper to discover the potential of politically relevant discussions or sequences in social media as a novel instrument to explore the extent of congruence between issue preferences of political elites and citizens. Our setting is formed by the planned six election debates broadcasted on TV and radio, leading up to the Dutch Parliamentary elections of September 12, 2012. The policy positions of party leaders will be captured by transcribing and coding the debates according to a predefined scheme. Furthermore, we will monitor citizens' attitudes on policy issues addressed by the candidates using Twitter messages sent during these debates. We will use software which mines Tweets containing a selected set of hashtags and assesses the sentiment expressed in them (Pang & Lee 2008). Through the simultaneous measurement of policy positions on both mass and elite side it will be possible to capture position distances on up-to-date issues around general elections. Moreover, applying the same measurement to six consecutive debates allows us to trace how issue discrepancies between citizens and parties develop on these topics in the last three weeks before the elections.

38. The Effect of Attempting to Recruit Respondents to a Web-Based Diary on Overall Response Rate

Michelle A. Cantave, *Arbitron, Inc.*; Robin Gentry, *Arbitron, Inc.*

Arbitron Inc., a provider of radio ratings data, conducted a test using a probability based address sample to recruit the general population, aged 13 and older, to complete a one week diary of their radio listening. Traditionally Arbitron uses a hybrid frame, which includes address matched and unmatched RDD sample, cell phone households, and no phone households, to recruit households for our one week diary. For the telephone numbers that

we have an address for: matched RDD, cell phone, and no phone households; we send a pre-alert letter before we call to attempt to recruit the household. Upon recruiting the household we sent a follow up letter, then the diaries, as well as follow-up phone calls for the households with phone numbers. In this test we attempted to recruit households from a phone matched address based sample using both our traditional recruitment practices (control group) as well as trying to recruit the household by sending them an invitation to complete the diary by going online (test group). For those households sent the online diary invitation, we followed up with nonresponding households and attempted to recruit them via our standard methodology. In this presentation, we will report the effects of first attempting to recruit the respondents to the Web based survey on the response rate (test vs. control group) as well as comparing the results from the phone matched address based sample to those recruited from the address matched RDD sample (control group vs. standard production).

39. Measuring Patient Health Behavior: Information Sharing With Healthcare Providers
Tammy J. Payton, *National Marrow Donor Program*; Heather K. Moore, *National Marrow Donor Program*; Jaime M. Preussler, *National Marrow Donor Program*; Viengneese Thao, *National Marrow Donor Program*; Michelle J. Kolb, *National Marrow Donor Program*; Navneet S. Majhail, *National Marrow Donor Program*; Elizabeth A. Murphy, *National Marrow Donor Program*; Ellen M. Denzen, *National Marrow Donor Program*

Bone marrow and cord blood transplant (transplant) is a potentially curative, but complex and resource-intensive therapy for patients with blood cancers as well as other genetic and immune disorders. It is estimated that there are currently 100,000 transplant survivors in the United States and this number is expected to grow two- to three-fold by 2020. Studies to date have shown that the quality of survivorship care is frequently suboptimal, and as a result, survivors are often lost to systematic follow-up within the healthcare system. The literature also suggests that the majority of cancer patients rarely or never discuss information they find important with their provider. As such, patient-focused post-transplant care guides were developed to facilitate follow-up care, especially the transition of care from transplant specialist to local physician, and to promote patient-provider information sharing. To evaluate the effectiveness of the guides overall, and specifically in addressing this issue, a longitudinal, repeat-measures survey was administered at 6, 12 and 24 months post-transplant to a nationally representative cohort of transplant recipients. The challenge was to measure patients' information sharing experiences in a single question with a focus on minimizing respondent burden. We will describe survey instrument pilot results and question design as well as characterize differences between patient groups who do and who do not share information with their providers. These results can be used to improve the precision of information sharing measures and identify communication barriers. Addressing these barriers may ultimately improve patient-provider decision making and patient satisfaction.

40. Using Focus Groups to Develop and Understand Survey Questions
Kinsey Gimbel, *Fors Marsh Group*; Katherine Ely, *Fors Marsh Group*; Bryan Wiggins, *Fors Marsh Group*; Jennifer Romano Bergstrom, *Fors Marsh Group*

Although often viewed as a qualitative data collection tool, focus groups can be a powerful tool in the survey development process. While cognitive interviewing is a more traditional way of testing survey questions, focus groups can also be structured so as to guide and assist in the development of both high-level survey topics and specific questions. This can be done before, during, and after survey development:

- Before work begins on survey design, focus groups can help researchers identify key concepts and topics to include in the study, and help spot subjects that will not be profitable survey areas.
- During survey development, focus groups can be used to evaluate prospective survey questions, identify possible response options, and refine question wording.
- After data collection is complete, focus groups can be used to better understand survey findings.

This paper will use examples from a series of focus group projects conducted for the Department of Defense during 2011 and 2012 to illustrate how focus groups can be used at each point in the survey process to improve survey materials and better understand survey findings. Areas of discussion will include specific questions and activities used during groups to solicit responses, examples of how questions were modified based on group feedback, and how focus group discussions can be used to expand on concepts used in survey questions.

41. Effects of Displaying Videos on Measurement in a Web Survey

Jonathan Mendelson, *Fors Marsh Group*; Jennifer L. Gibson, *Fors Marsh Group*; Jennifer Romano Bergstrom, *Fors Marsh Group*

Advertisers often use videos in online surveys to assess effectiveness of advertisements. While this allows marketers to test immediate reactions to videos, technical issues and lack of high-speed Internet access can introduce issues of generalizability and of comparability with alternate methodologies. Despite increased interest in embedding rich media in surveys, there is little published research on the implications for survey measurement. In a probability-based online advertising tracking survey, respondents were asked two sets of advertising recall questions. First, they were asked if they had seen advertisements for the Military or for any of its specific Services. Next, depending on whether respondents could successfully view a test video, respondents were shown videos or images of several specific advertisements and asked if they had seen them. Respondents who had seen the advertisements or who were shown videos were asked about their reactions to the ads. Time spent per survey page and the randomized presentation order of the advertisements was recorded. Our research examines the effects of using video stimuli on measurement. First, we use logistic regression to predict whether respondents could view videos, based on demographics; differences would indicate potential bias in studies solely using a video-based methodology. Second, we examine differences in ad recall based on whether respondents were shown images or videos, using demographics and the first set of recall questions to attempt to control for the possible confound between respondent selection into the video condition and respondent ability to view videos. Third, we use regression methods to predict whether respondents who were shown videos viewed the entire advertisements, based on demographics and presentation order. Fourth, we examine response differentiation and the selection of 'not sure' options in the ad reaction questions among respondents who were shown videos, based on demographics, whether respondents viewed the entire videos, and presentation order.

Saturday, May 18
1:15 p.m. – 2:15 p.m.
AAPOR Demonstration Session #3

Simulating the Effect of Follow-Up Survey Response Rates on Program Outcomes

Rebecca Lien, *Professional Data Analysts, Inc.*

Using data from three tobacco cessation phone counseling programs (quitlines), we simulate program outcomes at lower survey response rate levels than what was achieved. Program outcomes explored include quit status and program satisfaction measured seven months after program registration. The quitline field has adopted a target of 50% for follow-up surveys, yet the majority of U.S. quitlines do not achieve this target. We conducted the simulation as a tool to discuss the importance of survey response rates to the quitline community. We collected intake and 7-month follow-up data for quitlines in the three states: Minnesota (n=1,287); Florida (n=3,430); and Hawaii (n=1,203). The survey response rates for the three quitline case studies ranged from 48% to 64%. Using the number of days from the first survey attempt to the survey completion date we calculate response rates and outcome measures at each day of the survey period. We then graph the outcomes as a function of the calculated survey response rate to show what quit rates and satisfaction measures would be for the same group of participants at lower survey response rates. We find the quit rate outcome is influenced more by the survey response rate than the satisfaction outcome in the three case studies. The simulation is straightforward and generalizable to other fields with follow-up surveys. The graphs were a useful tool in discussing non-response bias to the quitline community.

A Demonstration of the University of Michigan Survey Research Center's Electronic Listing Program

Frost A. Hubbard, *Survey Research Center, University of Michigan*; Jennifer Kelley, *Survey Research Center, University of Michigan*; Jeffrey Smith, *Survey Research Center, University of Michigan*; Xuetao Zhang, *Survey Research Center, University of Michigan*

In 2006, the University of Michigan's Survey Research Center developed the Electronic Listing Program (ELP), which enables us to do traditional and depending listing electronically. As defined by Eckman (2010), dependent listing occurs when field staff are given a list of addresses for a specific geographic area and asked to update the list based on what they find in the area in person. Traditional listing occurs when no address list is given to the field staff in advance and the staff member must create the entire list of addresses in the area. Doing both types of listing electronically has greatly reduced our listing processing error and processing costs. Since the inception of the ELP, we have continually revised the software to achieve three objectives. First, we made it as easy as possible for our field staff to rearrange the addresses on the list and to put them in "walking-sort" order as defined by Kish (1965). Second, we improved the quality of our listed addresses to reduce returned mail, but also our ability to match our listed addresses to commercial databases from vendors such as Marketing Systems Group and Acxiom. We accomplished this by parsing the addresses into the seven unique fields as defined by the USPS (e.g. housing unit number, street suffix). Finally, to have a clear sense of how many addresses were added, deleted or modified by field staff during the listing procedure in each geographic area, for each individual address, the ELP now transmits indicators of whether

the address has been added, deleted, or modified to our master listing database. With these indicators, we have data which will help us more accurately predict the areas in the future where we can forego dependent listing and select addresses directly from the USPS Delivery Sequence File.

Demonstration of an Integrated Respondent Management and Data Collection Tool for Mixed-Mode (Phone/Web/Mail) Surveys

Harlan Luxenberg, *Professional Data Analysts, Inc.*; Julie Rainey, *Professional Data Analysts, Inc.*

Many research and evaluation firms are recognizing the importance of collecting data through multiple modes in order to increase response rates and reach a more diverse pool of respondents. Firms often use Microsoft Excel for managing contact lists across modes or a pricier CATI system solution which may or may not meet all of their needs. Neither of these options fit the needs for our evaluation organization so we built our own tool based on our experience and survey methodology. This demonstration will showcase Synchronized Survey™, a tool that Professional Data Analysts, Inc. developed specifically for mixed mode data collection and has been using for the last four years. This software provides a central management interface for managing contacts across modes, tools for sending emails, processing mail merges, updating contact lists, tracking attempts and response, and entering mail surveys through an automated, dual-data entry system. Telephone interviewers have access to a secure interface which allows them to use caller lists to select which cases to attempt by viewing the complete annotated call history. An easy to use form leads them through the survey. Non-respondents can be automatically flagged to receive an additional mode and removed from all modes once they complete by any one mode. In addition, this software integrates with LimeSurvey, an open-source online surveying tool, so online surveys can be created in LimeSurvey, but managed through Synchronized Survey. A comprehensive reporting system shows real-time response and other metrics necessary for tracking multiple surveys and surveyors. The system is built using ASP.net technology and a SQL Server database. Surveys require a certain amount of programming to meet the needs of each project. This demonstration will showcase a recent tri-mode survey conducted using Synchronized Survey and LimeSurvey so others can learn what a homegrown, Mixed Mode survey application looks like.

RDC-in-RDC: A New Approach to International Data Sharing

Stefan Bender, *Institute for Employment Research*; Daniela E. Hochfellner, *Institute for Employment Research at the University of Michigan*; Margaret Levenstein, *University of Michigan*

International and comparative analysis is often difficult given the existing restrictions on access to non-public micro data. In most cases researchers are required to undertake a costly research stay at a foreign RDC to access the necessary data. In order to improve data accessibility for international researchers, the Research Data Center of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) in Nuremberg and the Michigan Center on the Demography of Aging (MICDA) have launched a new initiative in international data sharing, RDC-in-RDC. The RDC-in-RDC enables access to restricted German social security data stored on a secure server in Nuremberg from designated institutions with comparable standards but other locations. This is the first time that confidential German micro data have been made accessible to researchers outside of Germany. Researchers can apply for working with data on individuals, households, and establishments. The data contain daily

information on the employment and unemployment history of the individuals, occupations and education, wages, and benefits, as well as job search activities and job training schemes as covered by the German social security system. In all data sources it is possible to link individuals and households to establishments. Furthermore, access is granted to IAB surveys which also can be linked to the administrative records of the respondents and metadata like data on interviewers or non-response. Such kind of data access is important for social science in many ways. Globalization requires research of transnational topics, such as economic crises, migration and health. Moreover, the various linkage possibilities can be used to gain new insights in survey methodology. The paper contains a brief description of the RDC-in-RDC concept and its technical implementation. It provides an overview of the available data sources regarding comparative research topics and research on survey methodology.

Saturday, May 18
2:15 p.m. – 3:45 p.m.
AAPOR Concurrent Session I

Response Rates and Data Quality in Multi-Mode Surveys

Changing Horses Midstream? Mode Supplement Quasi-Experiment and Response Rates

Rumel Mahmood, *Center for Survey Research*; Mary Ellen Colten, *Center for Survey Research*; Jack Fowler, *Center for Survey Research*; Carol Cosenza, *Center for Survey Research*

Declining response rates for Random Digit Dial (RDD) samples, the traditional workhorse in survey research, over the past few decades has led to some consternation among survey researchers, and as a result helpful suggestions to improve response rates. For a survey on rationing in Medicare and high health care costs carried out for the University of Pennsylvania Medical School by the Center for Survey Research at the University of Massachusetts- Boston, from May- August, 2012, we adopted many of these best practices at the outset for our telephone survey: implementing a multi-frame sample (n=2800) with RDD (800), list (1400), and cell phone (600) components; sending pre-notification letters to those respondents for whom addresses were available (1568); and including a small monetary incentive (\$2) with the advance letters. (Since our survey was on Medicare, rationing, and health care costs, we sought to speak with a member of the household over the age of 40.) Despite these measures, our response rate was lower than expected. We decided to send a printed questionnaire to respondents for whom we had addresses but were unable to reach over the telephone or who refused the telephone interview. With the paper instrument we sent a letter tailored to the non-response type and a further incentive (\$10). Of the initial 200 surveys we mailed to non-respondents, we received 122 completed surveys (61%). After such a high yield, we mailed a printed questionnaire to the remaining non-respondents in our sample. In total, we obtained 388 telephone interviews and 503 completed mail surveys, for a final response rate of 50% (AAPOR 4). In this paper we present differences in the characteristics of those who responded via the two modes and some of the substantive differences that resulted from adding the mail responses to those from the telephone interviews.

Differential Incentives in a Dual Mode Survey of Health Care Providers

Brian Roff, *Mathematica Policy Research*; Kirsten A. Barrett, *Mathematica Policy Research*

Health care providers respond to surveys at very low rates. Mail surveys are commonly used when surveying physicians and similar health care professionals. Increasingly, surveys are being administered by Web or by mail with a Web option. The Web offers an opportunity for data to be collected more efficiently – data entry costs are reduced, data quality is improved, and respondent burden is reduced. Prior research on the dual mode mail/Web approach has focused on response rates, with mixed results (Schneider et al. 2005; Friese et al. 2010; McFarlane et al., 2009). Little research exists on the role incentives play in mode choice, especially when the incentive favors a certain mode. The use of differential incentives in dual-mode mail/Web surveys to encourage Web response in particular has not been examined in the physician population, although it has been studied in surveys of recent college graduates (Mooney et al., 20012). Mathematica Policy Research conducted a dual-mode mail/Web survey of a nationally representative sample of 5,000 health care workers providing care to patients with HIV/AIDS. To control survey costs while at the same time encouraging response, we offered a \$20 pre-paid incentive and a differential post-pay incentive that favored Web survey completion. Those responding via mail received an additional \$20 while those responding via Web received an additional \$40. Since we did not have emails for sample members, precluding an email invitation, we hypothesized that 60 to 70 percent of the responding clinicians would complete the survey by mail. However, only one third did—two-thirds responded by Web. In this paper, we will: 1) explain the rationale for using a differential incentive as a means to encourage mode selection, 2) describe differences between Web and mail survey responders, and 3) provide suggestions for improving dual-mode surveys and incentive structures in the future.

Suppressing Survey Response: Further Evidence to Not Use Web Instruction Cards

Orin T. Puniello, *Bloustein Center for Survey Research, Rutgers University*; Marc D. Weiner, *Bloustein Center for Survey Research, Rutgers University*; Robert B. Noland, *Alan M. Voorhees Transportation Center*

By way of a survey research experiment, Messer and Dillman (2011) theorized that an illustrated, explanatory “Web card” would increase Web response rates when stimulating survey participation via postal mail. While those authors found no such effect, the Web cards in their experiment were generic for all respondents. As personalization of a survey invitation tends to increase response, we theorized that personalization of the Web card would increase its efficacy. We embedded an experiment in an Internet survey driven by address-based sampling mail contacting. The sample (N=8,000), geographically centered around eight train-stations, was divided into three categories: no Web card; generic Web card; and, personalized Web card, i.e., preprinted with the respondent’s Internet survey passcode. Hypothesizing no effect for the “no card” and “generic card” respondents, we anticipated a response rate boost for the personalized Web cards. We found no effect in the proportion of Internet and mail response; however, while we found no effect on overall survey response in the “no card” and “personalized card” categories, we found a noticeable response suppression effect in the “generic card” category (N=6,938; $\chi^2=4.74$, $p=0.029$). An inferential logit model controlled for 1) whether the invitation letter, per se, was personalized; 2) nature of the housing unit; and 3) geography. We

found, as now expected from the bivariate analysis, no effect from the personalized card (OR=1.01; p=0.858). All of the other controls were statistically significant and performed as expected (for e.g., for a personalized invitation letter, OR=2.65; p=0.000). The important empirical finding is that even under these controlled conditions, the generic Web card still suppressed survey response (OR=0.87; p=0.050). The instructive lessons for survey researchers are to not waste valuable survey resources on Web cards, whether personalized or not, and that there is a demonstrable risk that using Web cards may actually suppress survey response.

Approaches to Collecting Data Using Interactive Voice Response (IVR) for Address-Based Samples

Douglas Williams, *Westat*; David Cantor, *Westat*; Shannan Catalano, *Bureau of Justice Statistics*

Investigation concerning the use of Interactive Voice Response (IVR) as a data collection tool is not new. The advantages offered through IVR data collection are increased sense of privacy to encourage the reporting of sensitive behaviors, standardized interviewing, computer assistance to accommodate complex skip patterns, and reduced costs. For household surveys the traditional protocol for connecting respondents is for an interviewer to contact the respondent and transfer to the IVR system (Gribble et al., 2000). Concern with this approach is the potential for respondents to drop out during the transfer. This has been found to be as high as 30 percent (Tourangeau, 2004). The involvement of an interviewer can mitigate potential costs savings of an IVR, and the high drop off rate can counteract the reduced biases gained from increased privacy. The rise of address-based sampling approaches (ABS) affords the opportunity to invite participants through mail contact, maximizing cost efficiency and avoiding drop offs due to system transfers. The paper reports on the results of a field test conducted for the Bureau of Justice Statistics in 2012 which examined the feasibility of using IVR to administer the National Crime Victimization Survey (NCVS). The NCVS is a two-stage victimization survey which, in its present form, requires complex skip patterns that cannot be accommodated on a mail paper survey. In this test households were randomly assigned to either CATI Only, CATI with transfer to IVR (CATI-to-IVR), or Mail invitation to call the IVR system (IVR Only). This paper will compare the response rates from these different approaches, as well as the types of respondents that responded. Overall, the response rates for the IVR Only are equivalent to CATI Only and higher than CATI-to-IVR. The presentation will provide detail on these results, including characteristics of the respondents to each of the different modes.

AAPOR Updates: Reports From The Transparency Initiative and Non-Probability Task Force

This session will present a report on progress for two AAPOR initiatives: the Transparency Initiative and the Non-Probability Sampling Task Force. It will provide AAPOR members with an opportunity to engage in discussion and dialogue with members of these two groups.

Transparency Initiative Coordinating Committee Report
Timothy Johnson, *University of Illinois at Chicago*

Non-Probability Task Force Report
Reg Baker, *Market Strategies, Inc.*; J. Michael Brick, *Westat*

Social Attitudes: Race, Gender and Generations

Measuring Anti-Black Racism in the U.S.

Tobias H. Stark, *Stanford University*; Josh Pasek, *University of Michigan*; Trevor Tompson, *Associated Press-NORC Center for Public Affairs Research*; Jon A. Krosnick, *Stanford University*

Especially in the light of the recent election campaign of president Obama, interest among social scientists in racial prejudice remains as high as ever. However, four years after the election of the first Black president, we have not reached agreement on how survey researchers should measure racism. Some scientists prefer measuring racial stereotypes, others focus on affective measure of prejudice, some address the issue with implicit measures, and another group of scientists focus on measures of “new racism” such as symbolic racism. In fact, the field has evolved into camps that seem to doubt the validity of the others’ approaches. We try to build bridges across these camps by understanding the relations between the different types of racism measures. Multitrait-Multimethod models are applied to data from three recent representative U.S. national surveys that included the most commonly used measures of racism. We assess similarities and differences in how the measures associate with each other as well as with various predictors and outcomes of racism. We discuss advantages and limitations of the different racial prejudice measures and propose guidelines for future research on racism.

Integration and Segregation in 21st Century Schools: Voter Conflicts Over Equality, Local Control, and Community

Rachel L. Moskowitz, *Northwestern University*

This paper explores the competing meanings of equality, local control, and community for voters in the context of a local school referendum in Evanston, IL. In March 2012, residents voted on a ballot referendum that would levy taxes earmarked for building a new school in the 5th ward of Evanston. This ward is a historically black neighborhood that has not had a neighborhood school since racial integration of the school district in the late 1960s. Notions of equality and community control were at the heart of the Evanston referendum debate on building this new neighborhood school; providing equal access for all neighborhoods to local community schools was pitted against maintaining city-wide racial integration of schools. This original survey experiment of the election explores how important factors, such as race and group identity, affect individuals’ preferences for equality and community control both in the abstract and in these specific circumstances. The role information played in this preference formation is also seriously considered in this paper.

A Failure to Engage? An Examination of the Political Life of Generation X

Jon D. Miller, *International Center for the Advancement of Scientific Literacy*

In recent years and in recent campaigns, political analysts have asked whether the 80 million young adults that comprise Generation X have become or will become active participants in the American political system. Some journalistic characterizations of Generation X have painted them as “slackers” who are often disengaged with the political system – in contrast to the more

activist young adults who led the civil rights and anti-war movements of the 1960s and 1970s. The 26-year record of the Longitudinal Study of American Youth (LSAY) provides a strong empirical base for examining and testing the idea that most Generation X young adults (born between 1961 and 1981) have failed to engage with the political system. The LSAY is a national longitudinal study that was initiated in 1987 and continues to collect new information from the same 5,000 respondents each year. The participants in the LSAY represent the center of the age range for Generation X. Parallel to Jennings and Niemi's longitudinal study of high school seniors in 1965, the LSAY has collected a wide array of political socialization and participation data over the last 26 years. A comparison of the patterns found in these two studies will provide empirical evidence about the engagement of the young adults in Generation X and a comparison with the preceding generation of young Americans. Although data from the 2012 election are still being collected, the data from preceding decades will show that the level of political engagement by Generation X young adults has been higher than that of preceding generations. A set of two-group structural equation models will be used to validate this claim, but the results will be presented in a format that will be accessible to AAPOR attendees with and without prior training or experience with statistical models.

Framing the “War on Women”: A Survey Experiment on the Effects of Partisan Framing on Issue Perception and Vote Choice

Ashley A. Koning, *Rutgers, The State University of New Jersey*; David P. Redlawsk, *Rutgers, The State University of New Jersey*

Women voters were at the forefront of the 2012 election, and women's issues continually made headlines. These stories became part of an overarching assertion that a “war on women” was being waged. The Democratic Party originated the “war on women” frame to specifically attack Republican stances and legislation on reproductive health, contraception, and rape. The Republicans soon countered, however, by framing the “war on women” as an economic one. Republicans argued that the war was actually being waged by President Obama's administration, which caused women to suffer most in terms of jobs, unemployment, and poverty rates. The “war on women” thus became an enduring part of the campaign and a symbol for the battle over women voters. But which party had the more effective “war on women” frame? We know who won the election and who women voters favored, but how did these frames—the Democrats' health-based one and the Republicans' economic-based one—affect perceptions of the “war on women” and individuals' ultimate vote? This paper explores the “war on women” rhetoric by employing the two differing partisan frames through a survey experiment design. We test each frames' influence on whether voters perceived the “war on women” as real or myth, which party they thought was most responsible for waging it, and if it had any influence on voting. We argue that while voters and women overall will believe and be more influenced by the Democrats' frame, Republicans (and men) will show greater support for the “war on women” in their own partisan framing. This research follows the framing literature by showing how different frames can differently affect subsequent perceptions and opinions, as well as adds the assertion that partisans may be more susceptible to issues they would not traditionally support when framed within their own values and arguments.

Changes in Gender Beliefs in the U.S. from 1977 to 2010: Results from the General Social Surveys

**Duane Alwin, *Pennsylvania State University*; Paula Tufis, *University of Bucharest*;
Kristen Lee, *University of Buffalo***

This research examines secular change in gender beliefs from 1977 to 2010 using state-level GSS data. Processes of change in gender beliefs are found to vary across three historically relevant time periods and across segments of the population defined by religion, gender and region of the country. While there has been considerable growth across time in all groups in support of egalitarian gender beliefs, men tend to lag behind women in support of women's work roles. In a decomposition analysis, we find that the dramatic rate of intra-cohort change in beliefs reported from 1977 to 1985 declines in later periods for both women and men. Our findings are consistent with the claim that an anti-feminist backlash emerged in the mid-1980s and a period of stagnation in the growth of egalitarian beliefs predominated through the 1990s and the early 21st century. Both religion and region influence the nature gender beliefs, with distinct patterns being independently shown by both sets of factors. Regional differences reveal patterns consistent with state endorsement of the 1972 Equal Rights Amendment. Regional composition with respect to religious adherents accounts for some, but not all, of the differences between regions, and generally both religion and region contribute independently to levels of gender beliefs. There are very few statistical interactions between the components of secular change and regional and religious variation, suggesting that components of change throughout the periods studied are relatively immune to the level differences in beliefs due to regional and religious variation. Change components among women do not depend upon religion or regional categories. We conclude that analyzing change in different historical periods and geographic regions and within different segments of the population defined by gender and religion sheds new light on the processes of gender belief change in the U.S. since the 1970s.

Satisficing and Cognitive Shortcuts

The Relations Among Different Cognitive Shortcuts in Surveys

Roger Tourangeau, *Westat*; Rebecca Medway, *University of Maryland*; Stanley Presser, *University of Maryland*

This paper examines the issue of whether some respondents are consistently "bad" respondents, who use a variety of methods to get through a questionnaire quickly and provide data of dubious value. We examine a wide range of cognitive shortcuts, including choosing the first and last response options, yea-saying, giving don't know and no opinion responses, non-differentiation among answers to similar questions, reporting numerical answers as round values, and selecting status quo responses. Some of these are forms of survey satisficing but others are not. The data include responses from national face-to-face, telephone, and Web surveys. Across all three modes, we find little evidence that respondents who exhibit a high rate of shortcuts in the first half of a questionnaire also exhibit a high rate in the second half. In addition, we find weak correlations among the various forms of shortcutting. It could be that respondents have preferred modes for coping with the demands of survey questions, with some preferring DK responses, others non-differentiation, and still others yea-saying. Another possibility is that item characteristics (which affect how interesting and difficult an item is for different respondents) play a more important role in determining the level of shortcutting than respondent characteristics. A final possibility is that these shortcuts do not represent a single phenomenon, but are at best loosely related strategies for dealing with survey questions. We do

not find consistent relations between any respondent variables (such as educational attainment) and any of our measures of the use of shortcuts.

Mindful Responding to Questions: The Dangers of Survey Satisficing **David L. Vannette, *Stanford University*; Jon A. Krosnick, *Stanford University***

Respondent satisficing during surveys is a significant concern for researchers because of the implications for data quality. As such, considerable efforts have been made to measure, understand, and reduce survey satisficing since the early 1990s. Recently, promising new areas of research on survey satisficing have emerged; one of these is the application of the psychological concept of mindfulness to further understand the cognitive processes implicated when a respondent is satisficing. To achieve high levels of data quality, researchers strive to induce respondents to engage in optimal process for answering survey questions (Krosnick, 1991). This optimizing process refers to a respondent attending to the question at hand and then proceeding through the process of interpreting the meaning of the question, searching their memory for all relevant information, integrating that information into summary judgments, mapping those judgments into the required response format, and the reporting their response (Tourangeau, Rips, and Rasinski, 2000). Satisficing occurs when respondents deviate from this cognitively demanding process and provide answers that they deem to be satisfactory. Mindfulness during the survey response process refers to one possible mechanism through which respondents may be able to apply the mental control necessary to exert the considerable cognitive effort required to optimize their responses to the questions being asked. This is contrasted with mindlessness in survey responding where a respondent does not exert sufficient mental control to optimize the survey response process; this may be a pathway to satisficing behaviors and the associated low-quality responses to survey questions. In this paper, we seek to integrate the existing psychological research on mindfulness and survey satisficing to further develop our understanding of the implications of mindfulness and mindlessness for the survey response process. We will also make suggestions for best practices in survey design in order to elicit mindful responses to survey questions.

Effects of Respondent Reluctance, Mode, and Technical Difficulties on Straight-Lining and Refusals in a Mixed-Mode Survey **Jennifer L. Gibson, *Fors Marsh Group*; Jonathan Mendelson, *Fors Marsh Group***

This study extends past research examining the effect on data quality of experiencing technical difficulties with a survey. Past research finds that straight-lining, which can indicate satisficing, is predicted by respondent reluctance and mode. Given the popularity of mixed-mode surveys with a Web option, it is important to understand whether these method factors affect data quality. We examined straight-lining in a quality-of-life survey of military recruiters offered in Web and paper modes. Straight-lining was evaluated as a function of survey reluctance, mode, and whether a respondent experienced technical difficulties. Common difficulties were trouble logging into the survey and security restrictions on personal computers. Of the 3,957 participants, most responded via the Web (77%) and did not experience technical difficulties (95%). Moderated multiple regressions will be estimated to describe the association of survey reluctance, mode, and technical difficulties with three measures of satisficing behavior: straightlining, endorsing “n/a” or “don’t know,” and refusals. Results will indicate whether respondents taking more or less time to return a completed survey, using different modes, or encountering technical difficulties are more likely engage in different forms of satisficing. Interaction results will indicate

whether certain combinations (e.g., Web respondents who encounter technical difficulties and take longer to respond) exacerbate indicators of potential satisficing.

Use of Drag-and-Drop Rating Scales in Web Surveys and Its Effect on Survey Reports and Data Quality

Tanja Kunz, *Darmstadt University of Technology*

In Web surveys, rating scales measuring respondents' attitudes and behaviors by means of a series of related statements are commonly presented in grid formats. Besides benefits from using grid questions displaying multiple items neatly arranged and easy to complete on a single screen, grid formats often evoke satisficing behavior as respondents rush through a list of serial items quickly. This, in turn, might be at the expense of processing each item carefully resulting, amongst others, in less differentiated answers compared to using grids with fewer items or single-item per screen formats. The present experiment is designed to gain a better understanding of how respondents answer to rating scale questions and how the quality of rating scale answers can be influenced by different kinds of grid formats. For that purpose, two types of drag-and-drop rating scales are developed with the aim to retain the benefits of a grid format but preventing respondents from satisficing behaviors by either 1) dragging answer options horizontally arranged in the top row to the question items in the first column, or 2) dragging question items stacked in the top row to answer options in the first column. A 3 x 5 factorial design is implemented in a randomized field experimental Web survey conducted among university applicants (n=6000) with varying number of items (6, 10, and 16) presented in drag-and-drop formats or standard grids. Rating scale formats are examined in terms of response distribution and indicators of data quality (item nonresponse, nondifferentiation, acquiescence and extremity bias). Results indicate that while all rating scale formats yield comparable substantial responses drag-and-drop rating scales encourage higher item differentiation. However, results concerning other indicators of data quality are mixed which are discussed within the scope of the cognitive question-answer process.

MAPOR Student Paper Award Winner

Speeding and Non-Differentiation in Web Surveys: Evidence of Correlation and Strategies for Reduction

Chan Zhang, *University of Michigan*

The interactivity of the Web can be harnessed to improve online response quality. A small body of research has begun to explore interactive prompts to reduce respondent satisficing, i.e., providing adequate but not optimal answers. For example, in our earlier work speeding (responding very quickly), is reduced with an interactive, textual prompt when responses are very fast (< 1/3 second per word). These and other studies have focused on one satisficing behavior, although it's likely respondents who engage one satisficing behavior engage in other such behaviors while completing the questionnaire. In fact, emerging evidence suggests a strong correlation between two well-known satisficing behaviors in Web surveys—speeding and non-differentiation (giving very similar ratings in grid questions). Given that both speeding and non-differentiation are prominent satisficing behaviors, which one should be addressed through prompting and does prompting one behavior over the other differently impact data quality? We tested this in an experiment using a probability-based online panel. We compare two types of prompts in a series of grid questions, one targeting only speeding and the other only non-differentiation (we also include a control condition of no prompt). We find that prompting either

speeding or non-differentiation can curtail both behaviors on grid questions. This reflects the inherent correlation of these two satisficing behaviors, and more importantly, suggests that both prompts indeed lead to more thoughtful answers (in contrast to if the two types of prompts only had parallel effects where speeding prompts only reduced speeding and vice versa). In addition, both prompts seem to enhance the quality of answers other than grid questions, suggesting potentially broad effects on respondent performance. We will also report evidence about the impact of prompts on respondents' behaviors in subsequent surveys of this panel, and whether any carry-over effects differ between the two types of prompts.

Mode Choice, Respondent Engagement and Data Quality

Accessibility or Simplicity? How Respondents Engage With a Multiportal (Mobile, Tablet, Online) Methodology for Data Collection

Michael W. Link, *The Nielsen Company*; Jennie Lai, *The Nielsen Company*; Kelly Bristol, *The Nielsen Company*

While “choice” may be good for consumers, it is unclear whether or not mode “choice” helps or hurts in our efforts to collect data from respondents. Moreover, mobile technologies add a number of new dimensions to computer-assisted interviewing, including potential changes in location, the ability to communicate more readily with respondents (via triggered pop-up messages or IMS), and the potential to move from device to device throughout the day. We examine the impact of mode choice on respondents' willingness to keep a two-week activity diary. Utilizing a “multi-portal” approach (i.e., smartphone, tablet and traditional online), we selected approximately 400 respondents in two cities utilizing a dual-frame (landline/cellphone) sample. Respondents could provide their information throughout the day in any location using a smartphone, tablet, online access or any combination of these. Those without one or more of these devices in their homes was deemed “out of scope” for the study. The study highlights several important findings: 1) despite having access to multiple ways of entering information, the vast majority of respondents utilized only one; 2) traditional online access was the preferred mode of entry over mobile devices; and 3) there were significant differences in terms of age (over 50 years versus 50 years and under) in respondents' willingness (or ability) to use these electronic modes to keep a multi-week activity diary. The findings highlight many of the opportunities and challenges with utilizing some of the new technologies—singularly or in concert—as data collection modes.

Online Survey Participation via Mobile Devices: Findings From Seven Access Panel Studies

Michael Bosnjak, *GESIS-Leibniz Institute for the Social Science*; Teresio Poggio, *Free University of Bozen-Bolzano*; Frederik Funke, *LINK Institute*

The diffusion of mobile devices such as tablet computers and smartphones enabling respondents to participate in self-administered online surveys create new challenges for survey methodology in terms of measurement (e.g., equivalence of mobile versus traditional online instruments) and nonresponse issues (e.g., response patterns among mobile participants in comparison to desktop-based respondents). By merging available data from several online access panel studies conducted between March and May 2012 in Germany, we have addressed four nonresponse-related research questions. First, how large is the share of mobile participants when conducting online panel surveys overall? Second, how can the propensity to

choose mobile modes be explained? Third, do mobile participants differ on participation parameters, such as the number of completed questions, and the length of entries to open-ended questions? Fourth, does mobile participation change as more advanced technological features (such as Flash technology) are being embedded? The results to be presented show that 1) a considerable share of online panel members did participate using a mobile device, 2) that the propensity for choosing mobile devices to participate in online surveys is a function of age and gender (younger subjects and males are more likely to participate in this way, rather than older subjects and women), 3) mobile respondents did not substantially differ from traditional online survey respondents on an array of participation rate indicators. However, when using Flash technology, 4) mobile participants showed extraordinarily high dropout rates (about twice as much mobile drop-out rate when using Flash technology compared to traditional computers). Implications for survey methodology will be discussed, along with avenues for future research.

Mode Choice on an iPhone Increases Survey Data Quality

Frederick G. Conrad, *University of Michigan*; Michael F. Schober, *The New School for Social Research*; Chan Zhang, *University of Michigan*; Huiying G. Yan, *University of Michigan*; Lucas Vickers, *The New School for Social Research*; Michael Johnston, *AT&T*; Andrew G. Hupp, *University of Michigan*; Lloyd Hemingway, *University of Michigan*; Stefanie Fail, *The New School for Social Research*; Patrick Ehlen, *AT&T*; Christopher Antoun, *University of Michigan*

We now commonly choose the mode through which we communicate. For example, if immediate feedback is needed a phone call makes sense; otherwise, an email message is fine. Similarly, if a written record of the exchange is desirable, email or text is appropriate; otherwise, phone is better. Smartphones and tablets make mode choice particularly easy and routine: the options can be selected from a single device with one finger movement or voice command. Can this kind of mode choice add value to the survey enterprise by, for example, increasing respondents' commitment to the task when answering in a mode they have chosen? We conducted an experiment to explore how mode choice affects data quality, completion and satisfaction. 1268 iPhone users were contacted on their iPhones by either a human or automated interviewer via voice or SMS text. This created four modes: Human Voice, Human Text, Automated Voice, and Automated Text. In half of the initial contacts, respondents were able to choose their interview mode (which could be the contact mode); in the remaining half the mode was simply assigned. Overall, more than half the mode choices involved a mode switch. But just being able to choose (whether switching or not) improved data quality: when respondents chose the interview mode, there was less satisficing (rounded numerical answers and non-differentiation) than when the mode was assigned. There was a small loss of participants at the point the choice was made but those who began the interview in a mode they chose were more likely to complete it than respondents interviewed in an assigned mode. Finally, those who chose their interview mode were more satisfied with the experience than those who were interviewed in an assigned mode. The results point to clear benefits from mode choice and the importance of further exploration.

Comparing Tablet, Computer, and Smartphone Survey Administrations **Tom Wells, *The Nielsen Company*; Justin Bailey, *The NPD Group*; Michael W. Link, *The Nielsen Company***

“Survey respondents are increasingly attempting to take surveys on their mobile devices, whether researchers intend for this or not” (Cazes et al., 2011, p. 2). Approximately 50% of U.S. adults own a smartphone (Nielsen 2012; Smith 2012) and approximately 20% of U.S. adults own a tablet (Rainie 2012). These trends have serious implications for online surveys, especially for those not optimized for mobile devices. In this paper, we present results from tablet, computer, and smartphone administrations of a survey. Our main focus is on surveys taken with tablets and whether tablet survey administration is comparable to computer survey administration. There is currently very little research on tablet administration of online surveys, however, with tablet ownership on the rise, understanding the effects of this survey mode will become increasingly more important. In this study, we fielded a survey to a large, national sample of online panelists, who are also smartphone users. For the mode effect research being conducted, panelists were randomly assigned to a mobile app version or an online computer version of the survey. However, among the 711 respondents completing the online survey, 128 completed the survey using a smartphone mobile Web browser and 33 completed the survey using a tablet. We analyze three measures of survey taking behavior—breakoff rates, survey completion times, and item-missing data—among tablet respondents, computer respondents, and smartphone respondents (both mobile app and mobile Web respondents). Based on our analysis, tablet survey administration appears to be comparable to computer survey administration. Across each measure, differences in survey taking behaviors were small and not statistically significant. At the same time, with two of the measures—breakoff rates and survey completion time—we consistently uncovered differences between smartphone administration and computer administration, with differences being more pronounced among smartphone mobile Web respondents.

Mobile Browser Web Surveys: Testing Response Rates, Data Quality and Best Practices

Kyley McGeeney, *Gallup*; Jenny Marlal, *Gallup*

The rapidly changing technological landscape of the United States has important implications for survey researchers. The challenges are well known for outbound telephone surveys, but to a lesser degree for Web-based surveys. According to estimates by the Pew Internet and American Life project, 55% of cellphone owners access the Internet via their phone, and for many Americans a cellular device is their only Internet connection. Mobile devices provide instant connectivity, allowing respondents to take surveys at any time of day, no matter where they are located, which is an exciting prospect. However, very little research exists to date about surveys completed via smartphones and other mobile devices. It is unknown if surveys designed to be compatible with mobile Web browsers increase response rates, or if respondents who respond via a mobile browser are demographically different than desktop respondents. Further, it is unknown if best practices for the design of desktop based Web surveys translate to mobile based surveys. The present study was conducted using the Gallup Panel, a probability based panel of over 50,000 members who complete studies via the Web, mail, or telephone. Panel members were randomly assigned to one of 12 treatment groups that compare three different modes (traditional Web only, traditional Web plus mobile browser compatible and outbound), two treatment for length, and two treatments for question layout. Closed and open-ended questions were tested. Paradata, such as user agent string, time per

survey, breakoffs, and answer changes, were also recorded as part of the study. The results will be analyzed to better understand how mobile compatible surveys affect response rates, the representativeness of the sample, and data quality. The authors will draw conclusions about the costs and benefits of mobile compatible surveys, and make suggestions for best practices.

Research on Behavioral and Time-Use Diaries

Augmenting Paper Diaries With Phone and Web Data Retrieval: Is it Effective? Laurie Wargelin, *Abt SRBI*; Jason Minser, *Abt SRBI*; Zachary Homer, *Abt SRBI*; Anna Fleeman, *Abt SRBI*; Randal ZuWallack, *Abt SRBI*

From the 1960s to the 1990s, most Household Travel Surveys (HTS) were conducted entirely by self-administered pen and paper diaries sent via USPS mail. Starting in the 1990s and into present day, researchers have augmented the paper diaries with phone and Web technologies for HTS data retrieval. These electronic programs provide the advantages of offering sophisticated geocoding capabilities, in-program data checking, and monitoring for valid responses. Some researchers have speculated that the advent of advanced technologies will make the pen and paper retrieval method obsolete. However, since the introduction of multi-method retrieval options, only 15-25% of travel diaries have been completed by Web while recent evidence indicates that less than 25% of diaries are reported by phone. A majority of travel diaries are still returned by mail, as evidenced in the recently completed Metropolitan Council HTS (Greater Minneapolis), an interim report for the Southern California Association of Governments (SCAG) HTS Augment Survey, and the Pretest from the Delaware Valley Regional Planning Commission (DVRPC) HTS. This phenomenon may be explained by: 1) accessibility to electronic methods; 2) advanced modeling requirements have greatly increased respondent burden, making telephone-based reporting cumbersome; and/or 3) thoughtful development of paper diaries, relying on years of survey research, may prove more appealing to respondents. Our research will explore the variations in travel reporting for each retrieval method in three distinct regions of the United States – Northeast, Midwest, and West – and analyze any underlying socio-demographics related to retrieval method. In addition to documenting the socio-demographics by the three methods, this paper will explore the quality of data collected by each retrieval method. The findings provide great insight as to whether having options is effective and efficient for surveys.

Comparison of Instantaneous Mobile Time Use Data Collection Methods to Traditional Time Diary Methods Pat Graham, *GfK Knowledge Networks*

Time use studies frequently make use of recall time diaries, which require respondents to recall all of their activities for a period of time (usually the 24 hours of single day). While time diaries are considered a tried and true method for studying time use, there is ample literature documenting survey error and trade-offs with this approach (National Academy of Sciences 2000; Phipps & Vernon 2009; Robinson 1999). For example, time diaries elicit relatively low response rates that vary systematically along demographic lines, rely on recall information that is often incomplete, and are known to under-report secondary activities. One potential solution to these issues of data quality has been to make use of recent enhancements in the quality, management and technology of “mobile” surveys to collect several instantaneous measurements from respondents throughout the day. Respondents can be “pinged” at pre-set

times to record information about what they are doing, where they are, who they are with, and their thoughts and feelings. If they fail to respond to the first “ping,” then they can be reminded again with another “ping.” Surveys conducted through “mobile” devices, however, are not without their limitations, mostly related to screen size and usability. Moreover, nothing is known empirically about how frequently to “ping” respondents to maximize data quality. This study compares data collected over three 24-hour periods of time (including Super Bowl Sunday) using a traditional time diary recorded at the end of each day and instantaneous measurements made throughout the day using mobile technology. The two modes of data collection will be evaluated based upon the non-response, number of primary and secondary activities reported, number of individuals present with the respondent, completeness of responses and the concurrent validity between measurements. Within the “mobile” collection mode, we will also how number of “pings” impacts data quality.

Examining the Relationship Between Error and Behavior in the American Time Use Survey Using Audit Trail Paradata

Nicholas Ruther, *University of Nebraska – Lincoln*; Tarek Al Baghal, *University of Nebraska – Lincoln*; Adam Eck, *University of Nebraska – Lincoln*; Leonard C. Stuart, *University of Nebraska – Lincoln*; A. L. Phillips, *University of Nebraska – Lincoln*; Robert Belli, *University of Nebraska – Lincoln*; Leen-Kiat Soh, *University of Nebraska - Lincoln*

Audit trails, usage information produced during a computer assisted survey, are a form of paradata that allows researchers to examine how an instrument is used by interviewers or respondents in the course of an interview. This research uses audit trails and survey responses from the American Time Use Survey (ATUS) to examine the relationship between the audit trail paradata and potential errors in the ATUS. Previous research has examined the relationship to a much more limited source of paradata to issues such as data quality and survey breakoff (Gutierrez et al. 2011, Peytchev 2009). Research has also identified a number of potential errors in time diaries and specifically in the ATUS, such as missing key daily events (such as sleeping, eating, and grooming), providing consistently rounded answers to the duration of activities, and having memory gaps where some part of the recall period cannot be remembered (Fricker 2007, Phillips et al. 2012). The current set of audit trails paradata provide such useful but infrequently available data such as timing data (e.g. data entry timing, length of interview), key stroke data, the number of programmed prompts indicating data warnings, and how the data was reported and entered (such as using a precoded response option versus verbatim responses). These data are used, in combination with other potentially important variables such as indicators of cognitive ability and demographics, to predict the likelihood and amount of error observed in the ATUS using the various indicators. Initial findings show the importance of audit trail paradata in understanding error. For example, more verbatim entries used by a respondent are associated with higher rates of missing key daily events compared against pre-coded responses. Activity entry editing, on the other hand, is associated with less overall presence of this error, indicating a potential interviewer-respondent interaction in correcting errors.

What Are You Doing Now?: Audit Trails, Activity Level Responses and Error in the American Time Use Survey

Tarek Al Baghal, *University of Nebraska – Lincoln*; Lynn Phillips, *University of Nebraska – Lincoln*; Nicholas Ruther, *University of Nebraska – Lincoln*; Robert F. Belli, *University of Nebraska – Lincoln*; Leonard Stuart, *University of Nebraska – Lincoln*; Adam Eck, *University of Nebraska – Lincoln*; Leenkiat Soh, *University of Nebraska – Lincoln*

The American Time Use Survey (ATUS) is a time use diary where respondents report all activities they performed in a given day. The granular (activity level) data it provides sheds light not only on time use, but also potentially on memory and survey response processes. For example, activity level data identifies when in remembering the past day errors occur, which may assist in the study of memory structure and cues used for recall. Using a unique data set combining ATUS public use and audit trail data, this research examines activity level data to answer questions such as how people recall the length of time of different types of activities, how recall affects errors, and the impact of respondent level characteristics, (e.g. cognitive ability), on activity level reports. Initial results show that durations (e.g. doing an activity for 45 minutes) are reported for shorter activities, whereas start and stop times (e.g. completing an activity at 4 p.m.) are used for longer activities. Interestingly, the majority (76.9%) of reported gaps in memory was given as start and stop times, but errors of vagueness were more reported as durations (69.2%). Further, the majority (61.2%) of memory gaps occurred during “off-peak” hours, outside of the standard working hours of 9-5, whereas the reverse was true for vague reports; 77% of these errors occurred during standard work hours. The effect of respondent characteristics will be examined using hierarchical linear modeling. The results of this study shed light on memory and survey response processes, with implications for survey design, particularly for time diaries.

Troubles With Time-Use: Examining Potential Indicators of Error in the American Time Use Survey

Andrea Lynn Phillips, *University of Nebraska – Lincoln*; Tarek al Baghal, *University of Nebraska – Lincoln*; Robert Belli, *University of Nebraska – Lincoln*

This study explores six potential indicators of measurement error in the American Time Use Survey (ATUS), for the purpose of analyzing satisficing behavior in time-diary research. Possible reasons for satisficing behavior include respondents' busyness, their levels of social capital, their cognitive sophistication, and the difficulty of retrieving the information requested. This analysis builds on the research of Fricker (2007), who identified three “missing data” indicators of error in the ATUS: whether the respondent failed to report eating, sleeping, or “personal grooming” in the day in question. This paper conducts more detailed analysis of these indicators than has been previously done, and also examines an indicator of rounding of time spent on activities, an indicator of errors in travel reports, and the presence of “memory gaps” reported by respondents. Regression and structural equation modeling are used to identify the impact of demographic and other descriptive variables on error indicators. Direct and indirect effects of these variables on error indicators are found, but these effects are not consistent across indicators. For instance, hours worked and race are positively correlated with the likelihood of missing sleeping and missing eating, but are negatively correlated with missing grooming. Age, education, race, and sex are also found to have significant indirect impacts on the likelihood of rounding. In contrast to previous assumptions made in the literature, this study indicates that off-used error indicators in the ATUS do not measure a single latent construct of

satisficing behavior. However, cognitive ability and the difficulty of retrieval are identified as important factors influencing satisficing in the ATUS.

Mixed Topics in Questionnaire Design II

Determining Optimal Recall Period Length for Surveys of Payment Instrument Use in the Past

Marcin Hitczenko, *Federal Reserve Boston*

With the increasing ability to store and manipulate large amounts of information, we are increasingly learning about the world by gathering and analyzing data. While advances in technology have made it easier to collect this data accurately and often instantaneously, a great deal of research, especially in the social sciences, continues to rely on surveys. Much work has been done documenting that surveys often lead to inconsistent or erroneous responses. For this reason, it is fundamental to understand how the data collection process interacts with the cognitive process to affect the responses. In this work, we focus on the effect of the length of the recall period in surveys that ask individuals to aggregate past behavior for a specific timeframe. We limit ourselves to data collected by RAND and the Consumer Payment Research Center at the Boston Federal Reserve regarding reported number of uses of four different payment instruments within a year, month, week, and day of the survey. This data consistently shows that the average reported daily usage decreases as the length of the recall period increases. This well-known phenomenon introduces a tradeoff between the benefit of sampling more days and the potential bias introduced by memory decay as the recall period increases. We propose a general form for a stochastic model mapping the actual number of payment instrument uses to the value reported, as a function of the recall period length. We fit the models by utilizing data from the Diary of Consumer Payment Choice, also of the CPRC, that tracks individuals' payment behavior for three consecutive days. We then use the results to determine the optimal recall period length for each instrument, defined to be that which minimizes the mean-square error of estimates. Implications for other types of data are discussed.

Mechanisms of Reporting to Dependent Questions in Panel Surveys

Stephanie Eckman, *Institute for Employment Research*; Annette Jaeckle, *Institute for Employment Research*

Panel surveys are used to measure change over time, but previous research has shown that simply asking the same questions of the same respondents in repeated interviews leads to overreporting of change. With proactive dependent interviewing, responses from the previous interview are preloaded into the questionnaire, and respondents are reminded of this information before being asked about their current situation. Existing research has shown that dependent interviewing techniques can reduce spurious change in wave-to-wave reports and thus improve the quality of estimates from longitudinal data. However, the literature provides little guidance on how such questions should be worded. After reminding a respondent of her report in the last wave ("Last time we interviewed you, you said that you were not employed"), we might ask: "Is that still the case?"; "Has that changed?"; "Is that still the case or has that changed?"; or we might ask the original question again: "What is your current labour market activity?". In this study we present experimental evidence from a longitudinal telephone survey in Germany (n=1500) in which we experimentally manipulated the wording of the dependent questions and contrasted them with independent questions. We report differences in the

responses collected by the different question types. Due to the concern that respondents may falsely confirm previous information as still applying, leading to underreporting of change in dependent interviewing, we also test hypotheses about how respondents answer such questions. In these tests, we focus on the roles played by personality, deliberate misreporting to shorten the interview, least effort strategies and cognitive ability in the response process to dependent questions. The paper provides evidence-based guidance on questionnaire design for panel surveys.

Is Time on Our Side? Decomposing Survey Length on the Health and Retirement Study

Piotr Dworak, *ISR*; Heidi Guyer, *Institute for Social Research University of Michigan*

The effects of questionnaire length on respondent burden and response rates have been studied over the years. However, less attention is paid to what factors, other than content, may explain the variation in survey length and which factors have a positive, versus a negative impact on the interview experience. This analysis explores a rich set of paradata from the Health and Retirement Study to develop a more holistic view of the survey length and its impact on Respondent cooperation. The Health and Retirement Study (HRS) administers computer-assisted in-person and phone interviews to over 20,000 participants every two years. The questionnaire covers a wide range of topics and has grown in size and complexity since 1992. In 2010, the average interview length was 153 minutes for an in-person interview with physical measures and biomarkers and 86 minutes for interviews completed by telephone. Currently, the HRS survey length is analyzed using section-level timings but recent developments allow controlling for the objective length – the number of fields encountered during the interview. There is preliminary evidence that after controlling for the objective length, other factors related to respondent characteristics (age, gender, education, employment), interviewer characteristics (age, gender, performance, and tenure), and other characteristics related to study design affect the length of the interview. Based on the preliminary findings this analysis aims not only to discern the key predictors of the survey length but to estimate their relative contribution, which in turn may inform length-reduction initiatives and more refined data collection cost-models. In addition, capitalizing on the HRS longitudinal design, we will investigate the influence of the interview length on cross-wave participation.

Building a History: Collecting Comprehensive Employment Data in a Web-Based, Multi-Mode Survey

Melissa Cominole, *RTI International*; Chris Bennet, *RTI International*; Lesa Caves, *RTI International*

Event history analysis is an increasingly common technique used by social scientists to analyze change over time. Conducting such analyses often relies on the availability of historical information provided by survey respondents, for whom it may be challenging to recall events that occurred months or years in the past. As a result, when developing a survey that collects information conducive to such an event history format, there are several competing survey design priorities to consider. One goal may be to collect enough data to meet the analytic needs of a diverse set of data users. An additional goal may be to provide sufficient response options for respondents to easily and accurately convey their experiences over broad spans of time. Yet another goal may be to provide a suite of features (e.g., event history calendar, validations, cross-checks) that minimize recall error and encourage similar experiences across modes. Still

another goal may be to ensure that the survey is conducted as efficiently as possible in order to minimize the response burden for respondents. For this large, nationally representative longitudinal study of recent college graduates, it was necessary to balance these competing priorities when developing items designed to elicit a history of employment, unemployment, and job search activities in the four years after college graduation. Here we examine the impact of this balance using such metrics as survey timing data, item-level nonresponse, comparisons to responses from an earlier wave of the survey, and comparable estimates from benchmark data sources. We will offer suggestions for survey designers based on lessons learned during the design and implementation process.

Using Visual Design Theory to Improve Skip Instructions: An Experimental Test Nicole Gohring, *University of Nebraska – Lincoln*; Jolene Smyth, *University of Nebraska – Lincoln*

With the emergence of Address Based Sampling (ABS) and the availability of the Computerized Delivery Sequence File researchers are increasingly utilizing mail surveys. However, one drawback of the mail mode is that respondents have to navigate their own way through a mail survey without interviewer or computer assistance. Thus, a common challenge in questionnaire design is determining how best to provide skip instructions. Previous research has identified design strategies that decrease the frequency of skip errors, but even in the most effective treatments nearly 20 percent of respondents still make navigational errors (Redline et al. 2003). In this paper we report the results of a skip instruction experiment conducted in the 2012 Nebraska Annual Social Indicators Survey (NASIS; n=954; AAPOR RR1 = 27.2%). Two versions of the NASIS questionnaire were created drawing heavily on current visual design theory. The first contained conventionally designed skip instructions in which the response option that triggered the skip was followed by a right hand arrow and a verbal instruction to “Go to question #”. The second version also used a right hand arrow and identical verbal instruction on the response option that triggered the skip, but included up to three design alterations that we hypothesize will increase the effectiveness of the skip instruction. These included 1) the addition of a right hand arrow connecting the response options that did not trigger a skip to their follow-up questions, 2) indentation of the immediate follow-up questions to create hierarchical subgrouping, and 3) where necessary for 1 and 2, reordering the response options in the originating question. Preliminary results support our hypotheses in showing that Version 2 led to significant decreases in skip errors. In addition to reporting results, the paper will discuss best practices for designing skip instructions based on current evidence and visual design theory.

Panel Recruitment, Attrition and Data Quality II

After Your Interviewer Looks Under the Couch: Strategies for Handling Attrition in Twin Studies

Christopher Ojeda, *The Pennsylvania State University*; Veronica Roth, *The Pennsylvania State University*; Eric Plutzer, *The Pennsylvania State University*

Twin studies have proliferated in the social sciences, revealing that behaviors such as voter turnout (Fowler et al. 2008), general ideology (Alford et al. 2005), and many specific political attitudes (Hatemi et al. 2011) are heritable. These estimates of heritability are derived from the analysis of complex, and frequently longitudinal surveys, but almost never account for key features of survey design. Most important, twin analyses ignore the differential probability of

answering questions, thereby increasing the risk of biased estimates. We examine panel attrition as one potential source of bias in the estimation of genetic and environmental influences. Using the National Longitudinal Study of Adolescent Health data (Add Health), we consider if and how panel attrition affects the estimates of genetic and environmental influence on voting behavior. To do so, we proceed in two steps. First, we explain how attrition biases estimates in a twin study and then propose strategies for reducing the bias. Second, we demonstrate evidence of attrition in Add Health and then compare three methods for mitigating bias due to attrition: complete case analysis, inverse probability weighting, and multiple imputation. In our analyses, we use the first wave (N = 1,974 sibling pairs) and third wave (N = 1,456 sibling pairs) of Add Health, conducted in 1996-1997 and 2001-2002, respectively. Finally, we discuss the strengths and weaknesses of these methods and how each may impact estimates of voting behavior. We believe this study represents a critical first step in ensuring that biosocial models produce accurate estimates of political behaviors and attitudes, rather than estimates that may be artifacts of the data collection process.

Panel Attrition: Separating Stayers, Sleepers and Other Types of Drop-Out in an Internet Panel

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

Attrition is the process of respondents dropping out in a panel study. Errors resulting from attrition decrease statistical power and can potentially bias estimates derived from survey data. As panels are increasingly being used in the social sciences as a source of empirical data, a good understanding of the determinants and consequences of attrition is important for all social scientists who make use of panel study data. In many panel surveys, the process of attrition is more subtle than being either in or out of the study. Respondents often miss out on one or more waves, but might return after that. They start off responding infrequently, but participate more often later in the course of the study. Using current models, it is difficult to incorporate such non-monotone attrition patterns in analyses of attrition. Non-monotone attrition is common in long running panels, or panels that collect data frequently. In order to separate different groups of respondents that each follow a distinct process of attrition, a Latent Class model is used. This allows the separation of different groups of respondents, that each follow a different and distinct process of attrition. Using background characteristics for a panel survey of 8000 respondents who were recruited using a probability-based method into the Web-based LISS panel, I show that respondents who loyally participate in every wave (stayers) are for example older and more conscientious than attriters, while infrequent (lurkers) respondents are younger and less educated. We can link these characteristics to attrition theories, and show that our findings can be related to theories on panel participation and reasons for dropout. I conclude by showing how each class contributes to attrition bias on voting behavior, and discuss ways to use attrition models to improve the panel survey process

Panel Attrition and Weighting Adjustments for the ANES Time Series

Matthew DeBell, *Stanford University*

The American National Election Studies (ANES) Times Series surveys have been conducted during every presidential election since 1948 and are among the most widely used datasets in political science. The ANES interviews respondents before each election and interviews the same respondents post-election, with some losses due to attrition. Attrition in panels typically is not random and typically contributes to survey error. However, the ANES has never produced a

formal analysis of the effects of attrition on the Time Series sample, nor has ANES developed weights to adjust explicitly for attrition effects. In this paper we analyze attrition in the ANES 2008 Time Series study, assess the effects of that attrition on the accuracy of the survey's estimates, and implement weighting adjustments for attrition bias. We then assess post-adjustment accuracy and examine the effects of these adjustments on voter turnout and candidate choice models. Implications for adaptive design are considered, in which the quality of the post-election sample could be improved by targeting reinterview efforts on respondents whose likely attrition would be most harmful to the quality of the sample. We conclude with recommendations for procedures aimed at the prevention, measurement, and correction of ANES panel attrition bias in the future.

Retention and Attrition: A Comparison Across Ethnic Groups

Jennifer Parker, *RAND Corporation*; Kirsten Becker, *RAND Corporation*; Benjamin Karney, *UCLA*

In a longitudinal study on marital satisfaction, we focused on low income couples of differing ethnicities. Couples in which both partners identified as Hispanic or Latino were asked to sub-categorize themselves as Puerto Rican, Cuban/Cuban American, Dominican, Mexican/Mexican-American, Central American, South American, Other Latin American, Other Hispanic/Latino or Mixed Hispanic. Information was also gathered on the couples' ages, preferred language (English or Spanish), country of origin and parents' country of origin. We will explore differences in retention amongst Hispanic participants relating to Hispanic/Latino sub-categories, age, preferred language and country of origin. We will also compare retained couples to those not retained using the same demographic points, and discuss the reasons for attrition amongst couples not retained. Lastly, we will report preliminary findings in differences in retention and reasons for attrition between Hispanic/ Latino couples, African-American couples and white couples.

Re-Interview Bias in Panel Surveys: Results from a Seven-Wave Randomized Experiment

Sebastian Lundmark, *Gothenburg University*; Mikael Gilljam, *Gothenburg University*

Comparing panel samples and refreshment samples, previous studies have found significant re-interview effects on people's knowledge. Participation in previous panel waves tends to produce more knowledgeable respondents. However, in studies of people's beliefs, attitudes and voting intentions, only minor re-interview effects have been detected (Das, Toepel & Soest 2011; Lazarsfeld 1944). Most of these studies have used three or fewer panel-waves, and none of them have used a randomized experiment design. This study rectifies these shortcomings by using a seven-wave panel together with a randomized experiment design. With this different and more ambitious approach, we are able to study re-interview effects on beliefs, attitudes and voting intentions with a relatively large number of waves, and with randomized gaps. More specifically, the design consists of one group of respondents receiving five waves and two gaps, one group receiving six waves and one gap, and one group receiving all the seven waves. In addition, we also compare these groups with two refreshment samples of new panelists (one probability sample, and one non-probability sample, however not randomized). Preliminary findings show that the number of waves a respondent is subjected to affects their responses to questions on attitudes, beliefs and voting intentions. The results indicate that professionalized panelists and overestimated response stability are a non-negligible problem in panel surveys.

Sunday, May 19

8:30 a.m. – 10:00 a.m.

AAPOR Concurrent Session J

Reliability and Validity of Measurement

Parent and Teacher Ratings of Children's Approaches to Learning and Behavior: Do They Align and Are They Reliable?

Ashley Kopack Klein, *Mathematica Policy Research*; Lizabeth Malone, *Mathematica Policy Research*

Studies of young children often rely on indirect or proxy reports of children's behavior given the lack of direct measures and the costs of administration. To verify the data, multiple reporters are often asked to report on the same child. However, there are not clear standards for deciding which reporter to use when ratings vary and using all or randomly picking one reporter may confound measurement (Kraemer et al. 2003). Our study aims to answer three questions: 1) What is the reliability of parent, teacher, and assessor ratings of children's behavior? 2) How similar are parent, teacher, and assessor ratings? 3) How do reporter ratings compare to a direct measure of children's behavior? We use data from the Head Start Family and Child Experiences Survey (FACES) 2009 to answer these questions. FACES includes a nationally representative sample of 3,349 children and uses multiple methods to collect data on children from several sources. We focus on 1,000 children who entered Head Start at age four in fall 2009. Parent, teacher, and assessor reports of children's behaviors are used to construct two rating scales—approaches to learning/social skills and problem behaviors—which overlap with three domains of executive functioning: working memory, inhibitory control, and attention. We use Cronbach's alpha to examine the internal consistency reliability of the scales. We examine the correlations and net difference rate between parent, teacher, and assessor ratings and children's performance on the executive functioning Pencil Tapping Task (Smith-Donald et al. 2007). We explore differences by child and family characteristics. This study contributes to decisions surrounding the use of multiple reporters by comparing indirect ratings across reporters and linking them to a direct measure of the same construct. As surveys often have limited resources, we discuss the utility and/or added value of multiple reporters to measure children's behaviors.

Proxy Reports of Children's General Health Status and the Role of Reporting Bias in the Association Between Child and Maternal Health

Dana Garbarski, *University of Wisconsin-Madison*

Child health is an important covariate of a variety of individual and familial health and socioeconomic outcomes as well as an important outcome on its own. Given that mothers often report children's health status in large scale survey collection efforts, it is essential to gain an understanding of how and the extent to which mothers' and children's reports of children's general health status differ as well as account for the ways in which the association of child health with other familial outcomes of interest may be subject to the common method bias of being reported by the same person. Using data from the first wave of the National Longitudinal

Study of Youth 1997 cohort (ages 12 to 17) the analysis demonstrates moderate concordance between mothers' and children's reports of children's general health status. The analysis also demonstrates that additional measures of child health and sociodemographic covariates such as children's age, race or ethnicity, and household wealth have stronger relationships with mothers' compared to children's reports of children's general health status. Finally, it appears that maternal reporting bias may lead to overestimation of the relationship between child health and other maternal-reported outcomes. Using maternal health as the criterion of interest, this analysis incorporates interaction effects to examine whether the statistical effect of child health on maternal health is greater when child health is reported by the mother compared to when the child reports it. This method gives researchers some idea about how much their results may be influenced by the common method bias of being reported by the same person based on a few assumptions, and is easier to incorporate than some of the more complicated methods for dealing with common method biases.

Differences Between Self-Reported and Actual Income: An Analysis of Low-Income Households Seeking Housing Assistance

Ahuva Jacobowitz, NYC Department of Housing Preservation and Development

Elyzabeth Gaumer, NYC Department of Housing Preservation and Development

Socioeconomic status is a key predictor in a wide range of disciplines and research questions. Researchers often rely on self-report income data; however, the validity of self-report answers is less well understood. In particular, capturing accurate income data using self-report can pose a difficult challenge as it is often considered a sensitive topic by respondents. To assess the validity of self-report income, we will conduct a comparison of household and individual income across different modes of data collection as part of a larger survey effort of applicants to New York City affordable housing. The population applying to affordable housing in New York City, and therefore the population for survey participants, is a near-poor, working population. They do not always have a traditional source of income, but rather work multiple jobs, have seasonal employment, or are self-employed, leading to both difficulties in calculating an annual income and potential error in reporting. We collect income data from two sources. The first is self-report household and individual level income listed on a household's housing application. The second is household and individual level verified income using pay stubs, tax returns, and employer verification among other sources as part of the verification process to determine eligibility. We will compare these self-report data against the verified income and analyze for reporting bias (n=1,000). Furthermore, since this is part of a larger data collection effort of a self-administered questionnaire that asks about other household information, we will do further analysis to look at trends in reporting across other variables of interest including race, education, neighborhood, and household composition. Since an error in self-report income could mean the difference between being determined eligible or ineligible for an affordable housing unit, this analysis has the potential to impact policies and interventions to help individuals more accurately report their income.

Measurement Error in Diabetes Patient Profiles: Demographic Differences Between Diagnosed and Undiagnosed Diabetics in a Large Nationally Representative Sample of Adults 25-34

Anna Bellatorre, *University of Nebraska-Lincoln*; Patrick Habecker, *University of Nebraska-Lincoln*

A wide body of literature exists documenting the rise in obesity in the United States in the past two decades. However, relatively little attention has been paid to the rise in co-morbid conditions such as diabetes, particularly undiagnosed diabetes in young adults. Existing information from BRFSS records indicate that the number of states with rates of diagnosed diabetes for all adults exceeding 9% of the population increased from zero in 1990 to fifteen in 2010, however no information exists for undiagnosed diabetes prevalence over that same time period. Using a nationally representative sample of young adults aged 25-34 from the National Longitudinal Study of Adolescent Health (Add Health); we evaluate the measurement error in demographics related to diabetes among this cohort. Using this data, we find that 59.4% of diabetes cases are undiagnosed among this cohort. Moreover, we find that significant bias exists in estimates related to race, gender, and overall health despite equivalent utilization of healthcare and insurance coverage when diagnosed diabetes is used as a measure for diabetes as opposed to using glycated hemoglobin levels exceeding 6.5% per deciliter to measure diabetes prevalence. We seek to use this adjusted profile of what diabetes looks like in young adults to inform the medical community on how best to catch cases of diabetes that would otherwise go undetected if the current profile were used to diagnose diabetes in young adults. Further, we seek to use this data to inform large national studies utilizing hemoglobin A1C on the importance of preventing race and gendered non-response for biomarker data collection.

Who Has What Information About Others: Proxy Reporting, Knowledge and Willingness

Katherine R. Kenward, *Research Support Services, Inc.*; Alisu Schoua-Glusberg, *Research Support Services, Inc.*; Eleanor R. Gerber, *Research Support Services, Inc.*; Patricia L. Goerman, *U.S. Census Bureau*; Elizabeth M. Nichols, *U.S. Census Bureau*; Murrey G. Olmstead, *RTI International*

The U.S. Census and other surveys typically collect data from households by asking a single household respondent to provide information about others that live in the dwelling. This method of enumeration assumes that the household respondent can act as an accurate proxy for all other household members and that he or she is willing to share information about all household members. This paper explores the cognitive strategies that people use when they are unaware or uncertain of the information they are being asked to provide as proxies and the extent to which it is possible to determine the quality of proxy responses in an actual enumeration. We also explore the reported willingness and/or barriers that exist when reporting for others in the household, especially those unrelated to the proxy. To explore these issues, we use data from cognitive interviews conducted with Census Bureau questions asking respondents about alternate addresses where household members may live or stay, such as former addresses, seasonal homes, or relatives' homes. We report what respondents think about responding for themselves, their family members, and those living at the same address who are unrelated or only tenuously attached to the household. We also describe strategies that can be used to determine the likelihood that the data are accurate and complete; also we identify alternative data collection strategies that may be warranted for households that include roommates,

boarders, or tenuously attached household occupants. Finally, the implications of the findings for the U.S. Census and other household surveys will be discussed.

Polling and Political Attitudes

Payoff at the Polls: An Investment Theory of Internal Political Efficacy

Tim Vercellotti, *Western New England University*

Research has found that voting for a winning candidate increases one's feelings of external political efficacy (the sense that government is responsive to one's needs). But little is known about the relationship between other forms of political activism on behalf of a winning candidate and internal efficacy (the sense that one can have an effect on politics). This research seeks to address that gap in the literature by proposing and testing an investment theory of internal political efficacy. Political scientists have speculated that internal efficacy is psychologically grounded in an individual's self-esteem and ego, and is therefore relatively stable and difficult to alter. I hypothesize that forms of campaign activity that require a greater investment of oneself, such as volunteering for a campaign, attending a political event or events, or urging others to support a candidate, are more likely to achieve the difficult task of increasing one's sense of internal efficacy when a voter's preferred candidate wins. Activities that require less of a personal investment, such as voting for a winning candidate, are less likely to alter feelings of internal efficacy. I test these hypotheses using a panel survey of Massachusetts voters interviewed before and after the November 2010 election for governor, as well as American National Election Study data for the same period. Controlling for existing levels of internal efficacy before the election, I find that high-investment activities are associated with increased levels of internal efficacy after the election, while low-investment activities are not. I also find that this is true for supporters of winning and losing candidates, suggesting that it is participation, and not the outcome, that makes the difference. Still, these results suggest that one's sense of internal efficacy is less fixed than previously thought, and that internal efficacy may be subject to change under certain circumstances.

MAPOR Student Paper Award Winner

The Influence of Competing Identity Appeals on Voter Participation

Samara Klar, *Northwestern University*; Spencer Piston, *University of Michigan*

Political rhetoric frequently targets specific identity groups in order to garner support from group members. Each year, pollsters and researchers note important voting blocs that emerge from such group-based appeals. A particularly effective tactic for increasing a demographic group's participation is to instill group members with a sense of anger. However, demographics illustrate that Americans are more likely than ever to identify with more than one identity group at a time—and, often, these groups may align with competing sides of a policy debate. The effect of targeting two competing identity groups on an individual's political participation is yet unknown. We administer a unique survey experiment to illustrate that political rhetoric targeting two competing identities actually causes group members to decrease their political participation, particularly with respect to one important activity: donating money. The results have implications for how political rhetoric may affect participation among highly coveted voters.

The 2012 Election: A Different Kind of Country

**Gary Langer, *Langer Research Associates*; Julie Phelan, *Langer Research Associates*;
Greg Holyk, *Langer Research Associates*; Damla Ergun, *Langer Research Associates***

“Protest or transformation?” was the title of our AAPOR presentation on the 2008 presidential election. Four years later, pre-election surveys and exit poll results in the 2012 contest point in the latter direction, underscoring demographic and related attitudinal changes that hold out the prospect of fundamental and potentially long-term changes in the nation’s political equation. Using 2012 results and previous decades of ABC News/Washington Post surveys and network exit polls, we will present a portrait of the forces at play in the latest contest for the White House, exploring preferences in partisanship, ideology and the role of government; views of the competing candidates and their policies; and the demographic shifts that informed the vote. Elements of the race we’ll trace include Mitt Romney’s starting position as the least personally popular major-party candidate in data at least since 1984, Barack Obama’s largely successful framing in the summer season, Romney’s transformation after the first debate and his advance in mid-October assessments, followed by a resurgence for Obama as the race drew to its close. We’ll present data showing the pre-election contest, by two standards of measure, as the closest either since 1960 or since the dawn of probability-based pre-election polling in 1936. Substantive topics of discussion will include the role of the economy and of the candidates’ economic empathy, including regression modeling identifying the strongest predictors of vote preference. We’ll also discuss record-setting or record-matching levels of polarization among groups (including men vs. women; young voters vs. seniors; and racial, partisan and ideological groups); and we’ll compare national exit poll and pre-election poll results.

The Impact of Political Sponsorship on Response to Political Surveys

Roger Tourangeau, *Westat*; Hanyu Sun, *University of Maryland*; Stanley Presser, *University of Maryland*

This talk presents the results from three experiments, exploring when and how the organization identified as sponsoring a survey affects who cooperates with the survey and the answers they provide. In the first experiment, a sample of people registered to vote in Maryland was randomly assigned to one of three conditions; a survey about politics was identified as being done by 1) researchers at the University Maryland; 2) the Campus Republicans at the University of Maryland; or 3) the Campus Democrats at the University of Maryland. To our surprise, we observed neither the nonresponse bias nor measurement bias that we believe most survey researchers would have predicted. That is, registered Democrats, Republicans, and Independents responded at essentially the same rate to the three conditions and gave essentially the same answers across the conditions. (We conducted half the experiment using mailed questionnaires and half using telephone interviews.) It is possible that the University connection in all three conditions undercut the partisan cue, but it is also possible the conventional wisdom about this kind of effect might be in need of revision—a possibility supported by the fact that, so far as we know, there have been no prior experimental demonstrations of a political sponsorship effect in the U.S.. To fill this gap, we conducted two more experiments in the context of actual political polls done just prior to the 2012 election. In two state polls, conducted by telephone, half the cases were told that the poll was being done “on behalf of Democratic candidates” and the remaining cases were not told this. We should have the results in the next few weeks.

The Influence of Core Political Values on Attitudes Towards Contentious Science
Patrick Sturgis, *University of Southampton*; Nick Allum, *University of Essex*; Ian Brunton-Smith, *University of Surrey*

Science and technology (S+T) are increasingly entering the public sphere as politically contested phenomena. In the USA, partisanship is now an important predictor of attitudes towards stem cell research, global warming, evolution and other areas scientific research. In this paper we develop this line of research to consider the influence of left/right political orientation and libertarian/authoritarian values on a particularly contentious area of research: biotechnology and genomics. Using data from the British Social Attitudes Survey, we test the hypothesis that conservative economic values are associated with support for genomics research while social conservatism constrains support and that both aspects of political values condition the way that citizens select and deploy information that amplifies conflict. We present the results of our analysis and derive some conclusions about how citizens make judgments about S+T that are consistent with their existing political predispositions.

Cell Phone Samples: Coverage and Weighting

Finding the Optimal Allocation of Sample Sizes in Dual Frame RDD Telephone Surveys

Haci Akcin, *CDC/OSELS/PHSPO*; Denise Bradford, *Northrop Grumman*

Random-digit dialing (RDD) telephone surveys have long been used to capture data about a target population. To maintain survey coverage and validity, surveys have had to add cellular telephone households to their samples. The Behavioral Risk Factor Surveillance System (BRFSS), for example, one of the largest state-based RDD telephone surveys, began conducting a large pilot study to collect cell phone data in 2008. In 2011, landline and cell phone data were combined and released for public use. Optimal allocation of samples in dual-frame (cell and landline) telephone surveys, however, is still not well defined. In this study, we examined data from the 2011 BRFSS with different characteristics: landline only, combined data with current allocation, and combined data with proposed optimal allocation. The study determines whether there is a cost-effective and optimal sample design feasible for dual-frame RDD telephone surveys.

Attempting to Boost RDD Cell Sample Productivity by Identifying Non-Working Numbers Prior to Dialing

Missy Mosher, *SSI*; Jonathan Best, *Princeton Survey Research Associates International*

To mitigate rising coverage bias from cell-only households, telephone studies of the general population are including a significant cell phone component in their design. Federal law prohibits phone rooms from using predictive dialers to call cell phone sample. Consequently, data collection costs are high as interviewers spend significant time manually dialing non-working cellular numbers. These costs create a demand for wireless sample that is screened for non-working numbers before it reaches the interviewers. SSI, in conjunction with Neustar Information Services, has developed a method for identifying non-working numbers and numbers that are likely to be non-working in RDD cell samples. Starting with a randomly generated EPSEM wireless RDD sample, the numbers are matched against an extensive caller ID network where telephone activity levels are tracked. Numbers with low activity levels are

identified and can be excluded prior to dialing thus increasing the working phone rate of the sample. Specifics of this process will be discussed. Additionally, the authors will analyze the accuracy of the coding and if using it can increase phone room productivity. The extent of potential non-coverage bias introduced by excluding cell phone numbers with low activity levels will also be explored. The information provided is essential to researchers making an informed decision on whether to screen their wireless sample.

Modeling Phone Usage to Weight Dual Frame Samples

Kristie M. Healey, *ICF International*; William Robb, *ICF International*; Naomi Freedner-Maguire, *ICF International*; Kurt Peters, *ICF International*

The use of a dual frame design for telephone based surveys is increasing, and in some ways has become the new standard, due to the increasing use of cell phones and the corresponding decrease in land-line only households. With these designs, telephone numbers are sampled from two frames, one representing land line telephone numbers and one representing mobile telephone numbers. There is significant overlap of the two frames. That is, respondents who use both landlines and cell phones could potentially be selected through either frame. Proper weighting of the data takes this overlap into account. Combining data from the two samples without adjusting for frame overlap will result in biased estimates. To make such an adjustment, we need data on telephone usage to identify dual users—those that use both types of telephone service—in each response group. Ideally, it is best to find out during interviewing whether respondents are dual users, cell only, or landline only. This paper evaluates an option for making the weight adjustment for dual frames when self-reported phone usage is not available. We used demographics from an existing dual frame survey to model the probability of dual phone usage separately for landline and cell data. This model was then applied to a dual frame survey where self-reported information about phone usage was not available. We compared weighted estimates for key survey findings using two sets of weights: those that included no dual-frame adjustment and those that adjusted for predicted phone usage. Finally, we applied the same model to a third dual-frame study and compared estimates from three sets of weights: adjusted based on known phone usage, those adjusted based on modeled phone usage, and not adjusted at all.

Estimation and Prediction of the Landline and Cell-Phone Incidence for Local Areas

Stanislav Kolenikov, *Abt SRBI*; Randal ZuWallack, *Abt SRBI*

Researchers designing dual frame samples must determine how to optimally allocate the sample across frames. This requires accurate cost information and population proportions; the latter is the subject of this paper. Overestimation of the cell-only population will result in an allocation that unnecessarily increases the project cost; underestimation will result in an allocation that produces higher sampling variability. National and regional estimates have been released biannually since 2004 based on data from the National Health Interview Survey (NHIS) (Blumberg et al., 2012a). Researchers have developed small area estimation models to estimate sub-regional estimates (Battaglia et al., 2010; Blumberg et al., 2011 and 2012). A limitation of these estimates is the lag time of about 10-12 months after data collection, compounded by additional lead from several months to several years between the sample design and the field period (e.g. to accommodate OMB or a long term contract.) We advance the current research in the area of cell-only prediction utilizing an alternative small area

approach that combines demographic data and telecommunications trends—such as the total number of landline access points, cell phone subscriptions, and the number of ported numbers. A multinomial logistic regression is formulated on NHIS data, where the response variable is the (three-category) phone usage and the explanatory variables are based on the household demographics. The model coefficients are plugged to ACS data and state-level predictions are obtained. Finally, the joint generalized method of moments objective function is formulated as a quadratic form in the multinomial score equations and the discrepancies of the model prediction from FCC counts. Thus the model respects both the small area demographic profile and the administrative records. We demonstrate how the model based on NHIS 2009–2011 performs in predicting the usage rates in 2012, and provide our predictions for 2013.

Impact of Weighting Methods on Tobacco Use Estimates from a Dual-Frame RDD Survey

S. Sean Hu, *Centers for Disease Control and Prevention*; Burton Levine, *RTI International*; Shanta Dube, *Centers for Disease Control and Prevention*

Differences in estimates of tobacco use among adults have been observed among the major surveillance systems including National Health Interview Survey (NHIS), Behavioral Risk Factor Surveillance System (BRFSS), and National Adult Tobacco Survey (NATS). Sample variance is least likely to be the reason for these observed differences and therefore differences in estimates are likely due to bias. For RDD dual-frame telephone surveys such as the BRFSS and NATS, low response rates and differential response rates across subgroups may increase bias in estimates of population parameters. To reduce potential nonresponse bias in RDD dual frame surveys, poststratification is used, which constrains the sum of the weights to equal external population totals based on combinations of geography, phone usage, age category, gender, and race/ethnicity category. However, constraining the weights to this set of population distributions does not effectively compensate for nonresponse bias. The purpose of the current study is to explore the combinations of characteristics to constrain population totals in the NATS weighting procedure for effectively compensating for nonresponse bias. Using data from the 2009-2010 NATS, we identified the variables that are most correlated with current tobacco use and response propensity. Then, we use raking and model-based poststratification procedures to constrain the sum of the weights to distributions of these variables attained through external data sources. Using the NHIS as a benchmark, since it has a relatively high response rate, we compare the smoking rates nationally and by state to determine which combination of constraints results in the least bias.

Sampling, Response Propensity and Weighting

Consumer File Ancillary Data and Nonresponse Adjustment: Assessing the Consistency of Estimates Across Weighting Strategies

Josh Pasek, *University of Michigan*; Curtiss Cobb, *GfK Knowledge Networks*; J. Michael Dennis, *GfK Knowledge Networks*

The increasing availability of auxiliary data sources that can be linked to data at the household level provides survey researchers with new sources of information about sampled units. Data sources such as consumer file ancillary data, paradata, and even social media data allow practitioners to assess differential characteristics of respondents and nonrespondents. What is unclear, however, is how effectively each of these new sources of information can account for

differences between individuals who do and do not respond to our primary data collection efforts. The current study compares the results of weighting techniques using consumer file ancillary data with those of more traditional corrections. Using a unique dataset collected by GfK where consumer file ancillary data was appended to all households in an address-based sample, we explore point estimates and relations between variables under a variety of weighting techniques. Specifically, we compare raking to CPS marginals, propensity score weights to the CPS, propensity score weights to the sample using the ancillary data, and multiple imputation to the ancillary data as means to derive estimates and relations linking a number of political variables to one another. We discuss the assumptions behind each of these corrective techniques as well as the implications of the differences observed.

Improving Data Collection Procedures Using Prediction Methods

Julia Lee, *University of Michigan*

Data collection procedures that are implemented under a conventional survey design may incur differential nonresponse among subjects with different characteristics. This differential nonresponse could lead to biased survey inferences. Responsive design, an alternative design strategy, monitors and uses process data ('paradata') to alter the design during the course of data collection. The process data guides data collection decisions and prioritizes subjects meeting certain criteria to improve both survey cost efficiency and the representativeness of the respondent pool. Under the responsive design framework, this research describes a model-based strategy that combines prediction and balancing using benchmark information from a high quality survey to improve sampling and data collection of a 'current survey' consisting of multi-phase data collection. Models predicting sample characteristics from frame and contextual information are fitted to data from the benchmark survey (such as ACS), which shares the same frame and contextual information as the current survey of interest. The fitted models are used to predict sample characteristics for the 'current survey' to guide sampling decisions aimed at obtaining samples that better represent the targeted population. The proposed method is illustrated using two large government surveys, treating one as the benchmark survey and one as the 'current survey'. Analysis of the observed data from the benchmark and 'current survey' suggests that respondent distributions of the current survey are different from those of the benchmark. The results of the simulated survey using the proposed method obtains respondents that better represent the target population. In addition, the inferences based on the observed current survey have larger estimated standard errors than those based on the proposed strategy. This proposal provides a framework for a stochastic data collection strategy that aims to simultaneously attenuate nonresponse bias and increase inference precision, while maintaining the same budget and timeliness of a conventional survey.

Will Snowball Sampling Leave Your Data in the Cold?

Kristin Cavallaro, *SSI*

As online research becomes more integrated into the everyday methodology of industries across the board, we find the need to target for very specific groups of people. Whether we need to target people who use a specific brand of antiperspirant or those with a rare form of cancer, some rare populations can be almost impossible to find on an online access panel. While the use of additional sample sources increases the feasibility for some of these projects there are still valuable untapped resources that could make a world of difference in the success of a project. The great advantage we have in the struggle find these rare populations is that people with similar lifestyles or experiences tend to cluster—often sharing similar beliefs or

banding together based on a commonality such as a disease, an interest in the same model car, or alumni from the same college. The practice of “Snowball sampling” (identifying one person who fits the profile and asking that person to “spread the word” within their community) has been a technique criticized by some, who have feared it will introduce unacceptable biases. But with average project incidences continuing to fall, it may be time to take another look. SSI will conduct side-by-side tests to compare data from snowball sample to both online access panels and intercept sample. SSI will also test to find the optimal combination of sources and sample types (panel, snowball, river, etc.) yielding the most sound data available from an online sample frame. Topics for this test will include consumer goods, healthcare, known offline benchmarks and more. The findings will help researchers in all industries create methodologically sound sampling plans as they have in the past with the possible introduction of a broader reach made possible by the use of snowball sampling.

Difficulty in Capturing Minority Populations in RDD Survey Through a Landline Oversample

Timothy R. Sahr, *Ohio Colleges of Medicine Government Resource Center*; Bo Lu, *The Ohio State University*; Marcus Berzofsky, *RTI International*; Amy Ferketich, *The Ohio State University*; Jamie Ridenhour, *RTI International*; Thomas Duffy, *RTI International*

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. In this paper, 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).

Methodological Briefs: Questionnaire Design

How Open Are We to the Open-Ended Questions?

Saida Mamedova, *American Institutes for Research*

Political polling and other opinion related surveys literature has a large body of knowledge on open- vs. close-ended questions. These surveys are often telephone RDD or in-person interviews or, more recently, Web-based surveys. High item non-response has been one of the major reasons for surveys to avoid open-ended questions whenever possible. Even the open-ended questions that are limited to filling out a text field have been known to have high non-response. National Household Education Survey (NHES) in 2011 administered a mail survey field test with imbedded experiments on open-/close-ended questions. The respondents to the survey were first recruited by filling out a screener questionnaire. After an eligible child was selected from the information in the screener, a more extensive topical questionnaire was sent. The follow-up survey asked parents about their child’s education and the parental care and

family involvement in child's development. Imbedded in the design, there were questions which were asked in one form as an open-ended question and in another form as a close-ended question. The two forms were tested experimentally. One such question was on how many times a child was read to in the past week: one set of parents received an answer option in a write-in form and another set of parents received an answer option in the form of categories. In this paper, we will explore the response rates for these open-ended vs. close-ended option items. Our hypothesis is that the open-ended items are skipped more often than the close-ended items. We will use logistic regression to estimate the likelihood of response for one type of question vs. the other, controlling for other factors that may affect the response. This study will build up literature on the open- vs. close-ended questions as it relates to the mail household surveys.

Navigating Complexity in PAPI: Improving Questionnaire Comprehension on a Multi-National Media Trend Survey

Darby Steiger, *Gallup*; Kersten Weisbach, *Deutsche Welle*; Leah Ermarth, *Broadcasting Board of Governors*

Members of the Conference of International Broadcasters' Audience Research (CIBAR) developed a core media consumption questionnaire in 2010 to ensure consistent and accurate measurement of key performance indicators in the context of growing competition in local media markets and at the same time ever-tighter budgets for public broadcasting. Compared with the previous International Audience Research Program (IARP) questionnaire, the CIBAR core questionnaire was designed to be shorter and tighter, and hence, better suited to the changing research environment of declining response rates, growing interview costs and weary respondents. In 2012, Gallup conducted a redesign of the instrument to further refine the usability of the instrument for interviewers and data entry staff in the more than 50 countries where the survey is administered by paper and pencil face-to-face interviewing. This paper will present lessons learned from two companion efforts: 1) a qualitative and quantitative study conducted by Deutsche Welle to test and compare the CIBAR core with the former IARP questionnaire and 2) a review of navigational improvements made to the instrument in 2012 by Gallup that have addressed many of the challenges identified in the IARP and original CIBAR core questionnaires. The results of this study will shed light on key challenges in implementing face-to-face paper and pencil surveys that involve complex skip patterns, multiple response items, and recall items.

Measuring Happiness: Evaluating Life Satisfaction Versus the State of the World

Jason Husser, *Elon University*; Kenneth E. Fernandez, *Elon University*

The social scientific study of happiness has grown increasingly prominent. For instance, Federal Reserve Chairman Bernanke recently called on scholars to create better measures of well-being. We evaluate a common question designed to measure happiness: "Taken all together, how would you say things are these days--would you say that you are very happy, pretty happy, or not too happy?" Through two representative survey experiments, we show that the question is fundamentally flawed. Rather than measuring satisfaction with one's life, the oft-cited happiness question actually measures satisfaction with the state of the world, politically and economically. We suggest a simple correction of the question to better measure personal happiness.

Investigating the Effects of Questionnaire Design and Question Characteristics on Respondent Fatigue

Frida Vernersdotter, *The SOM Institute, University of Gothenburg*; Elias Markstedt, *The SOM Institute, University of Gothenburg*; Jonas Hägglund, *The SOM Institute, University of Gothenburg*

Overwhelming respondents with attitude. Looking for the contextual factors in questionnaires leading to breakoff. Survey noncompletion, breakoff, is often overlooked in the discussion on survey response rates as a proxy for data quality (Peytchev 2009). The contexts in which breakoffs occur have not been thoroughly investigated. In this study we investigate how the composition of question types affect breakoff propensity in the case of self-administered mail surveys. We examine the effects of questionnaire design, in particular frequency and concentration of attitude questions, on breakoffs. We draw on 26 years of consecutive self-administered mail surveys in Sweden, conducted by the SOM Institute at the University of Gothenburg, with a total of 73 000 respondents and 43 different questionnaires. The SOM surveys cover a wide range of topics in society, media and politics and are used for academic research on attitudes, values, self-reported behavior, and socio-economic status. The questionnaires are on average 22 pages long and have a mean response rate of 55 percent (RR1), 58 percent (RR2). For each of the questionnaires we identify the breakoff patterns in order to determine what questionnaire design and question features have caused them.

Investigating Signs of Interview Fatigue: Decreased Reporting of Category Expenditures

Brett E. McBride, *U.S. Bureau of Labor Statistics*

The survey design involving a screener or filter question followed by a series of more detailed questions is used in many surveys, including the Behavioral Risk Factor Surveillance Survey and the National Crime Victimization Survey. Some research has suggested that respondents learn that reporting a certain answer to a screener question will extend the interview through a series of follow-up questions and thus will alter their responses in a way that avoids the follow-up questions (Kessler et al., 1998). Additionally, a change in screener question response patterns over the course of an interview may reflect the cumulative cognitive burden that arises from a long interview. Whether due to respondent learning or fatigue, measurement error may be introduced into survey estimates. In the Consumer Expenditure Quarterly Interview Survey (CEQ), screener questions ask whether respondents have expenditures in various item categories over the course of an interview that lasts on average 56 minutes. Past research has found evidence of panel conditioning in responses to a screener question in one section of the CEQ (Shields & To, 2005). This research seeks to address whether there is a shift in responses to screener questions over the course of the interview and what may account for this pattern. The data examined comes from the wave one interview of the 2011 CEQ. Patterns of reporting expenditures are examined in the responses to screener questions asked of all respondents. In interviews involving a noticeable reduction in expenditure reporting, this research will identify whether measures indicating respondent reluctance or survey burden appear to be associated with the reduction. This paper will seek to disentangle the effects of decreased reporting and survey characteristics. Findings from this research will suggest whether new screener question formats or a reduction in survey length are warranted to confront decreased reporting of expenditure categories.

Measuring Issue Attitudes: Open Versus Closed Questions Redux

David RePass, *University of Connecticut*

For decades, social scientists and polling practitioners have debated the relative advantages and disadvantages of using open-ended versus fixed-choice questions to measure issue attitudes. In this paper, an extensive amount of survey data is examined in search of a definitive answer. First, let us postulate that if a person has an attitude, it will influence behavior. Indeed, many definitions of attitude include behavior as a component. This study tests the hypothesis that responses to fixed-choice issue questions are measuring issue attitudes and therefore should be related to voting behavior. Every one of the 208 issues asked in National Elections Studies since 1960 was correlated with vote (while controlling for attitudes toward the candidates and party identification). However, in only 17 of the 208 tests did issue position correlate significantly with vote. Thus, the null hypothesis was confirmed; fixed-choice issue questions do not measure attitudes. When the open-ended most important problem (MIP) question was tested in all elections since 1960, the issue attitudes ascertained by this measure were strongly related to vote, as strongly related to vote as party identification. Next, let us hypothesize that if a person has no attitude toward an issue, he or she will respond to a fixed-choice issue question in an inconsistent or random manner. The amount of such 'flip-flopping' can be observed by using panel studies. The National Elections Studies have conducted a number of panel studies over the past six decades. The author has developed a new measure that can estimate the amount of turnover in panel data. Using this measure, in 21 out of 25 fixed-choice issue questions asked in these panel studies, 56 to 77 percent of responses were inconsistent or random. The paper will also critique a number of studies that have examined fixed-choice versus open-ended methods of measuring issue attitudes.

Using Motivating Prompts to Increase Responses to Open-ended Questions in Mixed-mode Surveys: Where Should the Prompt Be Placed and to What Effect?

Glenn Israel, *University of Florida*

Getting respondents to provide high quality information to open-ended questions in self-administered surveys is a challenge. The evidence shows visual and verbal design elements play a role in response behavior. Regarding visual design, creating 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 information in the question stem was shown to improve 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 are used for the study. The importance prompt increased the item response rate for the question about improving Extension's services but it had no effect on the description question asking clients about getting information, its use and the result. In addition, the importance prompt increased the item response rate for mail surveys but not for Web surveys. 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 using a motivating prompt but it is unclear when and why it will be helpful.

The Influence of Answer Box Format, Personal Topic Interest, and Respondent Characteristics on Response Behavior in Open-ended Questions

Florian Keusch, *University of Michigan*

Previous research showed that the visual design of answer fields for open-ended questions in self-administered surveys influences response behavior depending on the type of response that is collected (Couper et al. 2011). For narrative responses, larger answer fields produce longer, more elaborated responses (Christian & Dillman 2004; Israel 2010; Stern et al., 2007), especially with less motivated respondents (Smyth et al., 2009). Questions that ask for frequencies and numeric responses seem to be less influenced by the answer space provided (Couper et al., 2011; Fuchs, 2009). Until now, no study has looked at the influence of the visual design of answer boxes in open-ended questions that ask respondents to list all known items of a specific category. Additionally, there is only limited research looking at the influence of personal topic interest on response behavior in open-ended questions (Holland & Christian, 2009). This paper looks at differences in response behavior (number of items named, item omission, response latency, and response order) between formats that provide the respondent with one large answer box or ten small answer boxes when asked for unaided brand awareness. In three experiments embedded in Web surveys, respondents from a non-probability online panel were randomly assigned to one of two question formats asking for unaided brand awareness of insurances (Experiment 1), airlines (Experiment 2), and car tires (Experiment 3). In two of the three experiments personal interest in the topic of the survey could be controlled for. The results of this study show that the number of brands named is significantly higher when ten small answer boxes are presented in two of the three studies indicating that respondents infer from the answer box format what the questionnaire designer expects from them. Personal topic interest and demographic characteristics of the respondents seem to play only a minor role.

International Public Opinion

The Americas Barometer: Public Opinion on Democracy and Governance Across the Western Hemisphere

Keith Neuman, *The Environics Institute for Survey Research*; Mitchell Seligson, *Vanderbilt University*

The Americas Barometer (www.AmericasBarometer.org) is a multi-country public opinion survey on democracy, governance and political engagement in the Americas, conducted every two years by a consortium of academic and think tank partners in the hemisphere under the general coordination of the Latin American Public Opinion Project (LAPOP) at Vanderbilt University. The Americas Barometer was first conducted in 11 countries in 2004, and most recently to 26 countries in 2012. It is the most expansive international survey project in the Western Hemisphere. In each country, the survey is conducted with a representative sample of voting-age adults, in all cases stratified by major regions in the country and in some cases including oversamples to provide for more in-depth analysis of groups (e.g., Afro-Colombians) or regions (e.g., internally displaced persons camps in Haiti). Surveys are conducted face-to-face with respondents in their households, except in the USA and Canada where surveys are conducted

online using established Internet panels. This research represents a unique body of public opinion data that is used extensively by academic researchers, governments, and organizations such as USAID, the World Bank, the Organization of American States, the Inter-American Development Bank and the United Nations Development Programme. The initial impetus for the Americas Barometer was to chart the evolution of democracy and civil institutions in Latin America and the Caribbean, but the issues covered are increasingly relevant to all countries faced with mounting challenges of governance, crime, corruption, political and civic engagement in the 21st century. This paper will introduce the Americas Barometer to AAPOR. It will provide a brief overview of this project as a unique case study of an ongoing multi-country collaborative project, and present selected findings from the 2012 survey with an emphasis U.S. public opinion in terms of trends over the decade and comparisons with Canada, Mexico, Latin America and the Caribbean.

When are Politicians Responsive to Public Opinion? Results from a Scenario-Based Survey of 3,000 Swedish Politicians **Patrik Öhberg, *Université de Montréal***

In representative democratic states, responsiveness is a core value. No matter how fine-tuned formal political rights or political institutions are, representative democracy does not function well without responsiveness. On a general level, the notion of responsiveness has to do with the connection between public opinion and public policy. Standpoints, priorities and values among voters are supposed to leave their mark on outputs from the political system. However, responsiveness is one of the most blurry notions within representative democratic theory and we need better tools to understand why politicians are responsive to public opinion in some situations, but not in other. In this paper, we try to contribute to the literature on responsiveness by asking politicians themselves under what circumstances policy decisions should be affected by shifts in public opinion. More specifically, this paper is the first to present the Panel of Politician conducted at the University of Gothenburg, Sweden, to an international audience. The panel includes almost 3000 politicians from local, regional and national levels. For example, 25 per cent of the country's MPs participate in the surveys. Given that Sweden has a little over 30 000 politicians, the number of participants in the panel is noteworthy. By presenting different scenarios where public opinion differs from the standpoint of the politician, we hope to identify mechanisms behind responsive behaviour. We vary the following mechanisms that can be assumed to affect responsiveness to public opinion: a) personal self-interest, b) policy area and c) different periods of the electoral cycle.

Social Media and Revolutions in Arab Nations: The Impact of Facebook on the Arab Spring **Muteb S. Alhamash, *Kingdom of Saudi Arabia***

It has been more than a year since the world watched the revolutions that shook the Middle East, the revolutions also known as the Arab Spring. There has been extensive material written about the internal factors (corruption, greed, nepotism, despotism) which led to the revolutions in Tunisia, Egypt, Yemen, Syria and Libya and there has been some material written about external factors. This paper explores the connection between the Arab countries that revolted and the use of social media sites, specifically Facebook, which acted as a "voice" for the people. It is hypothesized that Facebook had an impact on the revolutions, an impact that continues today. In addition to data from recent studies, this paper implements a survey which will attempt

to gather data from a pool of Arab citizens and will endeavor to understand the respondents' experiences with social media, revolution, and their perceptions of each. Key words: Arab Spring, Revolution, Social Media, Facebook, Tunisia, Libya, Egypt, Yemen, and Syria.

Interviewer Effects in the Arab Gulf: Lessons from Bahrain and Qatar

Justin Gengler, *Social and Economic Survey Research Institute, Qatar University*

Although the Arab world is experiencing a critical transition in the availability of systematic and objective public opinion data, researchers continue to rely on techniques developed in non-Arab societies to evaluate overall survey quality and estimate the total survey error. Interviewers are one of the sources of measurement error in surveys, and researchers have invested significant resources to create methods for detecting and reducing those errors. There are a handful of studies on interviewer effects in surveys conducted in the Middle East and North Africa, yet none examines how the ethnicity or nationality of an interviewer influences respondent answers to sensitive survey questions. Furthermore, no study of interviewer effects of any type has been conducted in the Gulf region, where the outwardly-observable categories of ethnicity and nationality retain special social and political salience. This study asks whether and why interviewer nationality and ethnicity affects responses to questions about political attitudes and behavior. Using data from the 2009 Arab Barometer survey conducted in Bahrain and two nationally-representative surveys conducted in Qatar in 2010 and 2013, the study finds strong evidence that the ethnicity and nationality of interviewers affects a variety of attitudinal questions related to sensitive social and political topics.

Freedom is in the Eye of the Beholder: Examining Perceptions of Media Freedom in China

Kay Ricci, *University of Nebraska – Lincoln*; Quan Zhou, *University of Nebraska – Lincoln*

Characterized by its stringent censorship practices and historical adherence to a “dominance model of media” (McQuail 2005), the government of China faces new challenges with respect to the Internet’s growing penetration of its population. According to the China Internet Network Information Center (CNNIC), Internet users in China have grown dramatically from 58 million in 2002 to 538 million in June 2012. Although the Chinese government has attempted to tighten its control over the Internet, this network remains relatively unrestricted when compared to other media. Thus, the Internet has fostered the rise of a public sphere that encourages interactions among its citizens (Yang 2003). This is in stark contrast to the traditional one-way communication in which the public only accepts the views disseminated by the government. Previous research has addressed the media’s effects on people’s confidence and trust of the political system (Chen and Shi 2001). This paper examines the Chinese public’s attitudes toward the media itself. Using data from the 2010 Gallup World Poll, a multinational probability-based survey, this paper examines the impact of critical factors, such as Internet access, education, confidence in institutions, and sector of employment, on the public’s perceptions of media freedom in China. Preliminary analyses suggest that a higher proportion of individuals whose homes lack Internet access believe that the Chinese media has “a lot of freedom” (75.9%), compared to those who report having Internet access (63.1%). Additionally, individuals with lower levels of education are more likely than those with higher education to think the media enjoys “a lot of freedom” (78.5% vs. 47.6%). Given that there is still a great deal of potential growth in Internet usage and changes in the educational system, our findings shed light on the development of China’s civil society and the changing attitudes of its people.

Investigating Challenges of Internet Surveys for Public Health Programs and Policies: From Neighborhood to Nation

The Triple Constraints of Health and Behavioral Surveys: Cost, Quality, and Time *Carol Crawford, Centers for Disease Control and Prevention*

Survey methodologists have always had to balance competing demands of lower costs, higher quality (coverage and non-response), and more timely data. The need to do so has become imperative and will continue to become more so in the face of austere budgets. Most door-to-door face-to-face surveys using multi-stage address-based samples, still considered the gold standard, gave way to random digit dialed (RDD) phone surveys because of cost and time. Now RDD phone surveys are facing considerable challenges. The population coverage rates are being eroded by wireless-only households, portable telephone numbers, telecommunication technology barriers (e.g., call-forwarding, call-blocking and pager connections), increased refusal rates and privacy concerns. While substantial research continues to alleviate many of these problems, the costs associated with RDD surveys remain high and the response rates are typically low. Moreover, the time from design to data release takes two years for most federal and state government surveys (e.g. National Health Interview Survey, Behavioral Risk Factor Surveillance System, and the California Health Interview Survey), making timeliness of the data less than optimal for efficient and effective public health programs and policy prioritization and evaluation. Different sampling frames, modes and analytical methods that may overcome these challenges and assist state public health professionals to continue to collect affordable quality and timely data that are representative of their respective populations are being evaluated. Novel approaches to health and behavioral surveillance include single and blended non-probability opt-in panels, and new statistical estimation methods. This presentation covers some of the novel approaches and preliminary results from pilot studies being conducted by the Division of Behavioral Surveillance, Centers for Disease Control and Prevention in wide-ranging public-private collaborations with states, academic researchers, and private companies.

Statistical Adjustments for Internet Opt-in Panel Surveys *Sunghee Lee, University of Michigan*

The data needs for producing population estimates for various subgroups at varying geographic levels in a timely manner are on the rise. Because it is difficult to satisfy those needs with traditional probability samples due to their high resource requirements, survey practice has turned to data collection using Internet opt-in panels. This practice, however, does not provide data with desirable unbiased properties due to the nonprobabilistic nature of the sample yet has outpaced the effort to understand the errors and to develop statistical methods to correct for them. In this study, we will use data from Centers for Disease Control and Prevention (CDC) funded study of health related quality of life and well-being measurement that included the ten-item measure from the Patient-Reported Outcomes Measurement Information System (PROMIS). These data were collected using an Internet opt-in sample that simulated census demographic distributions of the U.S. general population. PROMIS items, including the self-reported health item, have been part of well-established probability-based CDC national surveys. The analysis will focus on the comparison on these items across data sources. The comparison includes three types of statistics: 1) point estimates of the common variables for the

general population, 2) point estimates for the population subgroups (e.g., gender, age, race/ethnicity, education, geography), and 3) relationships across variables through regression modeling. The data will be used with and without weights in these comparisons. We will also discuss how such data may be blended with probability sample data. The findings from this study will enhance empirical evidence to understanding and using Internet opt-in panel data.

Internet Opt-In Panels Assessing Political Effects on Health Care **Stephen Ansolabehere, *Harvard University***

This paper analyzes standard measures of reported health effects from the 2012 Cooperative Congressional Election Study (CCES) and the 2012 NORC Election Study and compares them with survey results from prior national health surveys. The 2012 CCES consists of a 55,000 person sample of on-line respondents and was conducted by YouGov. The 2012 NORC study was conducted by NORC and consists of 2,000 random digit-dial phone responses from a mixed land-line and cell phone sample frame. I compare the national results from these two samples with prior national health surveys to gauge possible mode effects. I further study the variation on means, standard deviations, and correlations across states in the 2012 CCES. Prior research comparing on-line, phone, and mail surveys by Ansolabehere and Schaffner (2009) found no significant mode differences for health and political questions and demographics.

Identifying Sample Source of Sufficient Quantity, Availability, and Consistency to Meet Local Public Health Needs **Stephen Gittelman, *Marketing, Inc.***

There is an increasing interest in timely county- and community-level data to track health status, health behaviors, and health care access. Currently, the best reliable estimates come from aggregating three or more years of state health surveys using random digit dialed (RDD) dual-frame telephone surveys. RDD surveys face decreasing response rates, increasing costs, and infeasibility at county- and community-level. Surveying over 3,000 counties in the United States, would require a budget beyond that available in these difficult times. New survey and analytical methods that provide reliable estimates that meet local public health needs are needed. This study represents an initial effort at online data collection to address these challenges. The online double opt in panels have stood as the stalwart of sourcing for market research but the demands of the public health community for granularity and feasibility may outstrip the capabilities of current panels, estimated at 8 million members. A sample source of sufficient quantity, availability, and consistency has to be identified. Second, a criterion by which the sample frame is to be engaged has to be determined. Third, health related information is correlated to demography, those variables that are not so constrained must be identified and appropriate behavioral controls to balance the sample frame considered. No obvious covariance with the test variables needs to be demonstrated. This is an ongoing study in which preliminary data from four states will be available by the end of first quarter 2013. This presentation will present the results of this study to date and provide recommendations to support additional efforts in this area moving forward.

Cross-section vs. Panel Estimates of Vote Intention During an Election Campaign **Doug Rivers, *Stanford University and YouGov USA***

Analyzing change in voter preferences using repeated cross-sections depends upon stable sample composition. Researchers routinely weight samples to control for demographic variation, but have been reluctant to use attitudinal data for sample balancing, due to the lack of reliable benchmarks. In a panel design, however, it is feasible to correct for selection bias in multiple waves using baseline demographic and attitudinal data. Selection and weighting methods are described and evaluated using data from the 2012 U.S. elections.

Item Nonresponse: Prediction and Compensation

Predicting Item Nonresponse in a Recontact Study of Youth

Jennifer L. Gibson, *Fors Marsh Group*; Ashley A. Barbee, *Fors Marsh Group*; Luke Viera, *Fors Marsh Group*

Past research indicates that respondents may 'satisfice,' or conserve time and energy and yet produce an answer that seems good enough. This behavior, which is driven by respondent motivation, task difficulty, and cognitive ability, is likely to affect data completeness and quality and by extension the validity of study conclusions. Given the potential impact on study data, it is important to understand the various predictors of such behaviors when developing and analyzing survey results. The goal of this study is to examine predictors of two measures of data quality (item nonresponse and underreporting) for respondents of a recontact (advertising tracking) study of young adults who had completed a previous survey also regarding military recruiting. The advertising tracking survey follows an interleaved design. Each filter item indicating whether a respondent recalled seeing the target advertisement is followed by more detailed items only for respondents who answered a filter item affirmatively. Because respondents tend to learn that negative responses to filter questions help them complete the survey more quickly, some will begin to underreport recall as measured by the filter items. We examine underreporting and item nonresponse as functions of motivation and past behavior. Indicators of motivation assessed on the seed survey include demographic and attitudinal items related to interest in military service (i.e., relevance of the survey topic). Underreporting and item nonresponse on the advertising tracking survey are predicted based on these measures of motivation and item nonresponse on the seed survey.

Adjust Survey Response Distributions Using Multiple Imputation: A Simulation with External Validation

Frank C. Liu, *Institute of Political Science, National Sun Yat-Sen University*; Yu-Sung Su, *Department of Political Science, Tsinghua University*

One commonly acknowledged challenges in polls or surveys is item non-response, i.e., a significant proportion of respondents conceal their preferences about particular questions. This paper presents how multiple imputation (MI) techniques are applied to the reconstruction of vote choice distribution in telephone and face-to-face survey samples. Given previous studies about using this method in adjusting vote share information drawn from pre-election survey/poll data, this paper gives more attention to external validity of this method. Using survey data sets collected in Taiwan in early 2013, the authors take two steps to study the utilities of this method. First, they randomly take out a proportion (about 1/3 to 1/2) of values in a variable with few or no

missing values. Respondents of the missing values in the chosen variables are contacted. Their responses are compared against the “guesses” generated by MI. This paper reports and concludes the utility of applying MI to point-estimation adjustment.

Reduction of Item Nonresponse Bias by Accommodating Unequal Selection Probability in Multiple Imputation: Applications on Income Data in BRFSS and NHIS

Hanzhi Zhou, *Institute for Social Research, University of Michigan*

Income-related health inequality has been a special interest of researchers/agencies who conduct health surveys. However, the disproportionately high item nonresponse rates on income questions relative to other survey questions usually hinder such investigations. Although multiple imputation (MI) has been adopted by survey researchers to deal with missing data, there is inconsistency between the MI theory and its applications in practice. On one hand, data production in the public health and social science research is often based on complex sample surveys. On the other hand, existing software packages and procedures typically do not incorporate complex sample design features in the imputation process. Failure to account for design features can introduce severe bias on final estimates and hence invalid inference. In this paper, we applied a two-step MI method that was previously developed by us on two large public health surveys. Under the method, the complex feature of the survey design (including weights, clustering and stratification) is fully accounted for at the first step through a synthetic data generation procedure; conventional parametric MI for missing data is performed at the second step using readily available imputation software designed for an SRS sample. Data users need only to apply simple unweighted estimation methods to the imputed datasets. Using survey data from the Behavior Risk Factor Surveillance System (BRFSS) and National Health Interview Survey (NHIS), we evaluated the performance of our method in comparison with existing MI techniques. Extensive analyses are conducted on the income variable and related health measures for full-sample as well as domain estimation. The new method results in significant reduction in the bias, particularly in the presence of model misspecification or informative sampling.

Using Paradata, Questionnaire Characteristics and Respondent Characteristics to Examine Item Nonresponse

**Ana Lucia Cordova Cazar, *Gallup Research Center, University of Nebraska – Lincoln*;
Rebecca Powell J. Powell, *Gallup Research Center, University of Nebraska – Lincoln***

Because inaccurate data have little use, data accuracy is one of the main dimensions of survey quality (Biemer and Lyberg, 2003). Paradata, data about the data collection process, can shed light on ways to enhance data accuracy by allowing one to investigate factors that may lead to difficulties in the response process. When the response process is difficult for respondents, item nonresponse may occur. Item nonresponse is problematic not only because it has the potential to affect data accuracy, but also because it may create analytical difficulties as both effective sample size and statistical power are reduced (Beatty and Hermann, 2002). Cognitive processes that underlie a respondent’s decision to give an answer have received substantial attention (De Leeuw et al., 2003). The majority of these studies, however, have not used paradata to investigate item nonresponse. This study aims to fill that gap. Paradata and survey responses collected from the Internet component of the Gallup Panel are used to examine the extent to which characteristics such as the time spent filling out a questionnaire, the

questionnaire's topic, the respondent's level of interest in the survey, and the respondent's demographic characteristics influence whether the respondent completes the entire questionnaire. A two-part multivariate model will be used to predict whether a respondent gave an answer to every question in the survey, and if not, to identify the factors affecting the proportion of item nonresponse. In a sample of 17,045 respondents who answered a first questionnaire on media usage and a second questionnaire on world affairs two months later, preliminary analyses indicate that respondents' characteristics such as age and education are significant predictors of item nonresponse, and that these variables interact with survey topic and time devoted to answer the questionnaire.

Eliminate Item Non-Response: The Effect of Forcing Respondents to Answer in Web Surveys

Laura Leach, *Graduate Management Admission Council*

In Web-based surveys, a forced response to questions can be a solution to item non-response. This method may come with costs, however, that could affect the quality of responses and respondent drop-off rate. The Graduate Management Admission Council® (GMAC) conducted a Web-based survey with 4,135 motivated graduate business school alumni. GMAC investigated the impact that forced-response items had on respondent drop-off and qualitative differences in item answers. The survey used a random split-sample: Half of the respondents were forced to answer all survey questions. The other half was allowed to move to the next question without answering the current question. For the latter group, a request-response prompt notified respondents of an unanswered question and asked whether they would like to continue with or without answering the item. The survey was comprised of 49 questions and had an average completion time of 20 minutes. No differences were found between the forced-response and the request-response conditions with respect to respondent drop-off. In addition, no differences were found in the attitudinal nature of the response items. The forced-response and request-response designs had no impact on the response to categorical items regardless of placement in the survey. There was a marginal impact on items of personal sensitivity, such as compensation; however, this was not true for all finance-related questions. A motivated and interested population of graduate management alumni completed a lengthy questionnaire without regard to the treatment in this study. Furthermore, the content of responses was not impacted by forced or requested item conditions, and the only hesitancy was to reveal sensitive information, which is a common survey respondent issue.

Sunday, May 19

10:15 a.m. – 11:45 a.m.

AAPOR Concurrent Session K

Toward the Surveys of the Future

Envisioning the “Survey” of the Future: The Role of Smartphones and Tablets in Face-to-Face Interviewing

Robert Manchin, *Gallup Europe*; Femke De Keulenaer, *Gallup Europe*

The focus of this paper is on the role that technology can play in advancing the practice of face-to-face interviewing. More specifically, we will illustrate new ways of using smartphones and tablets during all stages of the data collection process, going beyond solely using these devices as a means to record respondents' responses. We will start by discussing how an application designed for use on a smartphone or tablet can help to construct an area sampling frame and draw a random sample. This application randomly selects one or more square-shaped areas (PSUs) in a municipality and samples a pre-defined number of points/locations in each PSU; each point is then “reverse geocoded” into an address and uploaded to the interviewer's device. At this point, the device becomes a direct assistant to the interviewer; e.g. interviewers can use built-in maps to navigate to the exact location/building that they have to locate. The second part will address issues related to collecting paradata and interviewer quality control. All aspects of the interviewer's task – locating addresses, completing contact forms, randomly selecting a respondent in each household etc. – can run via an application on the interviewer's smartphone or tablet. In other words, a large amount of paradata will be (automatically) collected and will be almost instantaneously accessible to the fieldwork managers, offering new possibilities for responsive survey design and interviewer quality control (e.g. via a built-in GPS locator and time stamps automatically attached to each step). In the final part of the paper, we will illustrate new ways to enrich survey data not only with location-related context data (e.g. using geo-location technology to link geo-spatial crime data to survey data), but also with “non-survey” data collected via the interviewer's smartphone or tablet (e.g. measurements of air quality).

Conversational Interaction and Survey Data Quality in SMS Text Interviews

Michael F. Schober, *The New School for Social Research*; Frederick G. Conrad, *University of Michigan*

Christopher Antoun, *University of Michigan*; Alison W. Bowers, *University of Michigan*; Andrew L. Hupp, *University of Michigan*; Huiying Yan, *University of Michigan*

As people increasingly adopt SMS text messaging for communicating in their daily lives, texting becomes a potentially important way to interact with survey respondents, who may expect that they can communicate with survey researchers as they communicate with others. Thus far our evidence from analyses of 642 iPhone interviews suggests that text interviewing can lead to higher quality data (less satisficing, more disclosure) than voice interviews on the same device, whether the questions are asked by an interviewer or an automated system. Respondents also report high satisfaction with text interviews, with many reporting that text is more convenient because they can continue with other activities while responding. Here we report analyses of how text interviews differed from voice interviews in our corpus. Text interviews took more than

twice as long, but the amount of time between turns (text messages) was large, and the total number of turns was two thirds as many as in voice interviews. As in our voice interviews, text interviews with human interviewers involved a small but significantly greater number of turns than text interviews with automated systems, not only because respondents engaged in small “talk” with human interviewers but because they requested clarification and help with the survey task more often than with the automated text interviewer. Respondents were more likely to type out full response options (as opposed to equally acceptable single character responses) with a human text interviewer. Analyses of the content and format of text interchanges compared to voice interchanges demonstrate both potential improvements in data quality and ease for respondents, but also pitfalls and challenges that a more asynchronous mode brings. The “anytime anywhere” qualities of text interviewing may reduce pressure to answer quickly, allowing respondents to answer more thoughtfully and to consult records even if they are mobile or multitasking.

Piloting a Mobile Data Collection Application: SurveyPulse™, by RTI International
David J. Roe, *RTI International*; Michael Keating, *RTI International*; Yuying Zhang, *RTI International*

The landscape of survey research continues to change with the evolution of mobile technologies and increased accessibility of Smartphones and tablet PCs. Both the adoption and the computing power of these devices are on the rise, providing users with increased exposure to information and opportunities to interact on a personal device. As a result, researchers must adapt to changing communication patterns and habits, and it is becoming more important than ever to explore the best methods for incorporating mobile data collection into survey research. While Smartphone survey applications (apps) have the potential to offer a robust set of features to researchers: instant data capture, real time insights, location data, multimedia access including video and the use of cameras, and better respondent communication tools such as push notifications, email and SMS (text), implementation, deciding what to implement and how to implement it can be a serious challenge. Many things must be taken into consideration, from building a custom app to buying into a panel using an already developed app, to data security, to the provision of user support. Further, applying best research practices for sampling, recruiting, coverage, and maintaining a panel of users must also be a part of exploration. This presentation focuses on the development and pilot testing of SurveyPulse™, by RTI International, from the decision to build a custom app, to recruiting and maintaining a panel of users. SurveyPulse™ is a mobile application designed to deliver surveys to users across multiple devices including tablets, platforms and operating systems and collect data in real time. Included is a discussion of app development and distribution, recruiting, data collection operations, data quality, user engagement and respondent communication. Also included is a discussion of plans for next steps, future research and expansion of this data capture method.

The iPad® Computer-Assisted Personal Interview system—A Revolution for In-Person Data Capture?
Heather Driscoll, *ICF International*; James Dayton, *ICF International*; Autumn Foushee, *ICF International*

In-person interviewing has long utilized paper-and-pencil surveys as the data collection mode for observational studies. At a time of increased scrutiny from the public and rising costs, electronic data collection devices are dramatically changing the landscape of these types of

studies. ICF has conducted several pilot studies using our iPad® Computer-Assisted Personal Interview system (iCAPI) since 2010 and found that it allowed for more efficient data collection, monitoring, cleaning, and analysis. Most recently, ICF conducted a study of the economic impact of Pennsylvania's water trails on the state's economy. This was our first complete implementation of our newly developed iCAPI. Our interviewers surveyed visitors to water trails (rivers that have been designated as a recreational water trail because they are important corridors between specific locations) at hundreds of boat and kayak launch sites during the summer of 2012. Through record heat waves, intense thunderstorms, and unpredictable site conditions, our interviewers successfully collected expenditure data from roughly 400 water trail visitors, using the iCAPI. Our most recent work in Pennsylvania confirmed and expanded on what we learned in our pilots, addressing questions, such as how easily can interviewers pick up the iCAPI system; how effective are the GPS and map capabilities; how do the iPads perform over weeks of data collection; and, is the iCAPI system greener, faster, better and cheaper? We were surprised by some in-field scenarios that were resolved with iCAPI; however, is the iCAPI system the perfect, sustainable intercept solution? Our paper will explore the advantages and limitations, as well as our ideas for refining the next iteration of applications for our iCAPI.

New Approaches to the Study of Attitude Formation and Political Behavior

A Multi-Survey, Multi-Methodological Assessment of Perception of Need and Quality of Life: Opinion Polling for the Common Good

Don Levy, *Siena Research Institute*

While public opinion polling is central to pre-election analysis, the sustainability of our craft may hinge on the degree to which we contribute to ongoing efforts to promote and enhance the common good. Locally, it may matter more to citizens to garner a clear understanding of shared need than what the projected vote totals may be in the next election. This paper discusses a methodological triangulation study measuring the perception of need in one northeastern county that includes a major urban center as well as variation in respondent quality of life and a ranking of governmental services. We conducted three surveys, two RDD with cells of the general public and one via mail/phone and Web from among service providers across non-profits, educators, public officials and clergy. Survey questions included multiple quality of life indicators, perceptions of need across multiple areas, and opinion questions pertaining to root causes of enduring societal problems and appropriate collective future directions. Data from the three surveys – 623 respondents to the Quality of Life survey, 1306 to the Community Needs Assessment and 391 to the multi-methodological service provider survey – were analytically merged with available secondary data and presented to the public through not only a report but also through publication in a local newspaper and a video on YouTube but also in three well-advertised public forums attended by over 200 residents. Using multiple surveys and methods, the variation in the public's perception of life in the county, the impact on some of their social location and the quality of local services was measured and reported to officials, service providers and the public. By making the data available in multiple forms and actively inviting comment and interactive discussion, the research stimulated collective response including the formation of an information and capacity sharing cooperative among local non-profits.

The Storm of the Century: Assessing the Effects of a Natural Disaster on Electoral Behavior and Attitudes

Krista Jenkins, *Fairleigh Dickinson University*; Dan Cassino, *Fairleigh Dickinson University*; Peter Woolley, *Fairleigh Dickinson University*

In October 2012, Hurricane Sandy hit the eastern seaboard and brought to an abrupt end attempts to conduct a pre-election survey in the days leading up to the presidential election. Almost two million New Jerseyans were without power, thousands were displaced, and telephone service (both cell and landline) was rendered inoperable for a large proportion of households. Prudence dictated that interviewing be suspended as even those residents who might have been reachable via phone struggled to recover from the storm. In short, the widespread nature of non-coverage was an insurmountable challenge to ongoing pre-election polling. However, rather than abandon the survey, our research design morphed into a panel study, whereby we recontacted the 400+ registered voters who were interviewed in the days preceding the hurricane's arrival. When power and phone services were widely restored we resumed the study and emerged with a unique data set from which to assess individual level effects of a natural disaster on electoral behavior and attitudes. Thus we revisited questions concerning an individual's voting intention, candidate preferences for both president and U.S. Senate, public questions that touched on referenda for higher education bonds, judicial fringe benefits, and favorability of key national and state political actors. The data and paper address several questions including "Given the opportunity to look presidential in non-partisan settings, do natural disasters increase the prospects for incumbent presidents?", "Do natural disasters heighten, diminish, or have no effect on one's likelihood of voting?", and "Are attitudinal and behavioral changes dependent on the degree of loss one experiences as the result of a natural disaster?" These questions, although basic, are rarely addressed given the infrequency with which natural disasters are so closely timed with elections.

Bayesian Estimation and the 2012 Presidential Election Exit Poll

Clint W. Stevenson, *Edison Research*

Election exit polling provides a unique opportunity to collect vote results as well as other information on the voting population immediately after a voter casts their vote on Election Day. Due to the nature of elections there is a significant amount of prior information available for each state, county, and precinct. This provides an excellent opportunity to apply Bayesian estimation to the exit poll data. Traditionally, exit polling is analyzed using frequentist approaches (e.g. hypothesis testing). This paper will discuss Bayesian approaches and how exit poll data can be analyzed and updated beginning at the start of Election Day until polls close. After polling locations close they often make the actual vote count available. All of these data (including all prior knowledge) can be combined to develop a Bayesian model to estimate the Election Night results quickly and accurately.

Preference-Based Measures of Media Exposure

Thomas J. Leeper, *Aarhus University*

Media exposure is among the most important constructs in political behavioral research, yet agreement on operationalization is lacking. Beyond susceptibilities to various biases, standard measures of exposure gloss over differences in informational content received by individuals who report similar levels of exposure. Prominent alternative approaches propose to measure

exposure through specific news stories or particular news programs. For use in most research, however, both approaches are burdensome on researchers and respondents and lack robustness across temporal, political, and media environments. Analyses of nationally representative Pew Research Center surveys from 1996 to 2008 and a large, online panel survey indicate that a preference-based measure of news-following offers a viable alternative that is more robust, strongly predicts variations in all extant exposure metrics, and is reliable at the individual and aggregate levels. These findings have implications for the conceptualization and measurement of media exposure and normative implications related to citizen awareness.

Separating Political Attitude Change from Attitude Uncertainty: (In)Consistency Experiments of the ESS Panel Component

Sedef Turper, *University of Twente*; Kees Aarts, *University of Twente*; Minna van Gerven, *University of Twente*

As far as its vital role for explaining causal mechanisms is concerned, change has always been a great interest to scholars. Scholarly attention paid tracing and explaining changes in attitudes and behavioral patterns of diverse populations, paved the way to many wide scale cross-sectional time-series data collection projects in the field of social sciences. However, while repeated cross-sectional surveys provide data about aggregate level trends, the evidence they provide about micro-level processes underlying these macro changes is indirect. Thus, the knowledge that standard cross-sectional studies can provide is destined to be incomplete in the absence of more direct evidence about micro-processes. This paper attempts to shed light on the micro-level political attitude change processes through (in)consistency confrontation experiments conducted as a part of the Panel Component of European Social Survey. In these experiments, a subset of the panel respondents are asked to confront with their responses from the previous wave, irrespective of whether they offered a consistent or an inconsistent answer. The design of the experiments allows us not only to systematically analyze the micro-level processes underlying political attitude change, but also to differentiate between genuine attitude change and attitude uncertainty. We first present to what extent level of attitude uncertainty and susceptibility to attitude change differ with respect to level of education, political interest, and attitude strength by using four-wave panel data representative of Dutch population over age 16. Secondly, we further investigate the nature of observed political attitude change among different education level, political interest and attitude strength groups through the examination of (in)consistency experiments. Analysis of the experimental data provides us with better understanding of attitude change at the micro-level and also with direct evidence needed to complement the statistical inferences on separation of attitude change from measurement error.

Investigating the Effectiveness of Incentives

Interviewer Attitudes and the Effectiveness of Monetary Incentives

Ulrich Krieger, *German Internet Panel*

Studies have shown interviewer characteristics, such as race, ethnicity, and gender can have a negative effect on data quality (Singer, Frankel, & Glassman, 1983; Catania, et al., 1996; Davis et al., 2010; O'Muircheartaigh & Campanelli, 1998). While much is known about how interviewer characteristics affect a survey respondent's answers, little is known about the measurement error effect due to interviewer attitudes on the survey topic. The known studies that investigate interviewer attitudes focus on attitudes towards their job satisfaction and performance and the

effect it has on production rates or data quality (Singer, Frankel, & Glassman, 1983; Hox, de Leeuw, & Kreft, 1991). There are no known studies that examine if interviewer attitudes on the survey topic have an impact on survey respondent's answers (i.e. measurement error). Using data from the National Survey of Family Growth, a national face-to-face survey that has both interviewer-administered (CAPI) and self-administered (ACASI) components, and interviewer characteristic data (i.e. demographics and attitudes), this study examines discrepancies between respondent answers from CAPI to ACASI on sensitive items (e.g. on number of sexual partners and abortions) and how interviewer attitudes on sexual behaviors and other demographics (e.g. age, religion) may relate to those discrepancies. The main hypothesis guiding this investigation is that interviewer attitudes about the survey topic, particularly sensitive topics, might unwittingly be transmitted to respondents and influence respondent answers for sensitive questions. Preliminary analysis shows that interviewer attitudes on sexual behaviors are correlated with respondent answer discrepancies from CAPI to ACASI, for both number of lifetime sexual partners and number of abortions. Further investigation is warranted and those results will also be reported.

The Influence of Respondent Incentives on Item Nonresponse and Measurement Error in a Web Survey

Barbara Felderer, *Institute for Employment Research*; Frauke Kreuter, *University of Maryland JPSM & IAB*; Joachim Winter, *University of Munich*

Even though a sampled person may agree to participate in a survey, she may not provide answers to all of the questions asked or might not answer questions correctly. This may lead to seriously biased estimates. It is well known that incentives can effectively be used to decrease unit nonresponse. The question we are analyzing here is whether incentives are able to decrease item nonresponse and measurement error as well. To study the effect of incentives on item nonresponse and measurement error, an experiment was conducted with participants of a Web survey. In addition to an incentive for participation, an extra prepaid incentive ranging from 0.50 Euro to 4.50 Euro was given to some respondents towards the end of the questionnaire in the form of an Amazon-voucher. At the same time, respondents were requested to think hard about the answers to the next questions and be as precise as possible. In this experiment there are two reference groups: one group received the request but no incentive and the other did not receive any request or incentive. The questions within the incentive experiment contain knowledge questions, recall questions referring to different time periods, and questions about subjective expectations. We approach our research questions in three steps: Our first analysis focuses on the effect of incentives on the proportion of "don't know"s and "no answer"s. In a second step, we look at the amount of rounding and heaping as an indicator for measurement error. In the third step, we examine measurement error directly for two variables (income, unemployment benefit reciprocity) by linking the survey data to German administrative records and computing the difference between survey response and administrative records. Comparisons across the different incentive groups will allow for an assessment of the effectiveness of incentives on item nonresponse and measurement error.

Improving Panel Maintenance Success on a Longitudinal Study

Tiffany L. Mattox, *RTI International*; Jennifer L. Domico, *RTI International*; Daniel J. Pratt, *RTI International*

Minimizing sample member attrition is vital to the success of longitudinal research. Key steps in this effort include periodically locating the sample members and confirming or updating their

contact information. RTI International is conducting the third follow-up survey for the Education Longitudinal Study of 2002 (ELS:2002), conducted for the National Center for Education Statistics, U.S. Department of Education. The study follows high school students over time to determine how their high school experiences influence their lives as they continue on to postsecondary education, the work force, and forming families. Sample members were originally surveyed as 10th grade students in 2002 (base year) and/or as 12th grade students in 2004 (first follow-up). The second follow-up was conducted in 2006. Panel maintenance activities were then performed at multiple points prior to the third follow-up. Given that the previous follow-up interview was conducted 6 years prior, we anticipated challenges in locating sample members for the third follow-up full scale data collection in 2012. Thus, we conducted an experiment with the third follow-up field test sample to determine whether offering \$10 to sample members – if the sample member or a parent updated or confirmed contact information on file for the ELS:2002 sample member – would increase panel-maintenance participation. The significant positive outcome of the experiment led us to extend this \$10 panel-maintenance-participation offer to the entire full-scale sample during the panel maintenance effort prior to the start of third follow-up full-scale data collection. In this paper we provide results from the field test panel-maintenance experiment and examine the panel-maintenance response from the full-scale sample prior to the third follow-up full-scale data collection. In addition, we examine the third follow-up full-scale survey response status of the panel maintenance respondents to gauge the ultimate success of these efforts.

50 Years Later: Do Respondents Who Remember the Initial Survey Provide Higher Quality Responses to a Follow-Up Survey?

Danielle K. Battle, *American Institutes for Research*; Rebecca Medway, *American Institutes for Research*

Groves, Presser, and Dipko (2004) found that people predisposed to be interested in a particular survey topic were more likely to participate in a survey on that topic. Studies focusing on topic interest have looked at its effect on cooperation with a survey request, but little research has evaluated the effect of topic interest on response quality. We hypothesize that those for whom the topic is highly salient are more highly engaged and thus put more effort into responding to survey questions. This paper presents results from the 2011-12 Project Talent Follow-up Pilot Study, which assesses the feasibility of reengaging a representative random subsample of the initial 1960 Project Talent participants. The initial Project Talent survey was a large-scale longitudinal study that collected extensive cognitive, personality and background information from 440,000 9th-12th graders in 1960. In the 2011-12 follow-up, participants were asked if they remember participating in the 1960 Project Talent study (about 60% did); we use the response to this item as a measure of topic interest. The follow-up also included a prepaid incentive experiment where participants were randomly assigned to receive no incentive, \$2, or \$20. This paper examines recall of the initial 1960 Project Talent study among the 2011-12 Follow-Up Pilot Study respondents and determines whether recall is predictive of response quality. It also looks at whether offering an incentive reduces any differences in response quality between those who do and do not recall the initial study. Response quality outcomes include item nonresponse, amount of time spent completing the questionnaire, straight-lining/non-differentiation response, round values, and consistency of responses to personality measures across the 1960 and 2011 collections.

Aspiring for More than Crumbs: The Impact of Incentives on Girl Scout Response Rates

Debra Dodson, *Girl Scout Research Institute, Girl Scouts of the USA*; Meredith Reid Sarkees, *Girl Scout Research Institute, Girl Scouts of the USA*; Cathy VonFange, *Abt/SRBI*

Youth development organizations are increasingly wedged between the demands of funders, on one hand, who want empirical evidence of program effectiveness and, on the other hand, a society increasingly unwilling to provide that empirical evidence – particularly when the data sought are from minors. This challenge requires navigating not only typical response rate challenges faced in survey research but also the additional complication of gaining parental consent in order to even contact those members under the age of 13. This paper draws on a summer 2012 study of Girl Scout members in 10 local councils to assess the relative effectiveness of a variety of strategies (virtual vs. traditional incentives; membership oriented vs. non-membership oriented incentives; and small rewards vs. chances to win larger prizes). The analysis explores the effectiveness of those incentives on willingness of parents to register girls, willingness of girls to respond, and the impact of incentives on representativeness of the respondents. The results can help us better understand the strategies for increasing accuracy of the data used to drive data-driven philanthropy.

Assessing Data Quality

Assessing the Quality of Survey Data Through Streamlined Data Processing
**Donsig Jang, *Mathematica Policy Research*; Amy Beyler, *Mathematica Policy Research*;
Alicia Haelen, *Mathematica Policy Research*; Flora F. Lan, *National Center for Science and Engineering Statistics (NCSES)***

Federal statistical agencies are continuously striving to provide high-quality survey data in a timely manner. Adaptive survey design (Groves and Heeringa 2006) is one method they are using to help achieve this goal. This type of design draws on several data sources, such as paradata, frame data, and processing data, in real time to help staff allocate resources effectively during data collection and make informed decisions about the closeout. The technological advancements that make adaptive survey design possible also make it possible to streamline data processing. Survey-management systems can now link data sources in real time, allowing statisticians to conduct editing, imputation, and weighting during data collection. Researchers can even monitor key survey variables during data collection. (These measures, along with R-indicators and response rates, can serve as indicators of survey bias.) Combining adaptive survey design with this streamlined process not only allows us to assess data quality and bias during data collection, but it also expedites data processing because it enables us to put all data-processing systems in place by the end of the collection period. The development of this process was motivated by the National Science Foundation (NSF). In conducting the National Survey of Recent College Graduates for NSF, we replaced the customary sequential approach to data processing with this integrated approach. This allowed us to test our data-processing procedures, including key SAS programs for autocoding, computer edits, and imputation. We produced and examined real-time quality measures, bias indicators, and paradata, and then assembled a comprehensive quality profile and assessed nonresponse bias. Monitoring the data enabled us to correct problems as they arose. We will present our data-processing framework, the measures we monitored during data collection, and the benefits and challenges of adopting this process.

Toward a Standard Toolkit for Comparing Samples: Point Estimates, Relations Between Variables and Trends Over Time

Josh Pasek, *University of Michigan*

The proliferation of both new methods for collecting data and novel analytical tools for translating between respondents and the population present exciting possibilities for public opinion research. But for researchers interested in understanding the population, these new opportunities may be accompanied with inferential pitfalls. Researchers need to identify the circumstances under which non-probability surveys, corporate data, and social media data can yield valuable insights and when these sources might instead lead to erroneous conclusions. Similarly, corrective tools such as raking, calibration, and matching have the potential to ameliorate some sources of survey error, but may be unable to adjust for other systematic biases. For survey researchers to fully utilize diverse sources of data to make conclusions about the population, they need to be able to assess how the conclusions from diverse data sources compare to one-another. Particularly, we need to know the circumstances under which the conclusions reached from these newer tools mirror those of more traditional analyses. In this paper I present a new toolkit for comparing the inferences derived from different sources of data and weighting strategies. Programmed as a freely available R package, the toolkit represents a standardized system for comparing the inferences derived from different datasets regarding point estimates, relations between variables, and trends over time. To illustrate the features this new software, the paper presents the results of a novel analysis of 16 weeks of comparable data collected from one probability RDD telephone data stream and one opt-in non-probability Internet data stream collected in the run-up to the 2004 U.S. Presidential election. The results show both the potential for a standardized comparison toolkit as well as the differences that can be observed across differing types of inferences.

Controlling Survey Response Bias with Range Regression Techniques

John Tuhao Chen, *Bowling Green State University*; Yuanting Zhang, *U.S. Food and Drug Administration*

Response bias arises when the respondent provides inaccurate information, possibly due to a leading survey question or social desirability bias. There is a lack of innovative methodologies that systematically deal with response bias. In this paper, we propose a new method called the range regression to analyze a dataset containing several waves of Health and Diet surveys (HDS) conducted by the U.S. Food and Drug Administration between 1982 and 2008. Range regression recently emerged in studying vascular surgery procedures regarding the amount of treated clots and post-thrombotic syndrome for patients with deep vein thrombosis. Intrinsically, range regression consists of stratification of respondents with similar ranges, followed by identification of a measure that bundles subject variability within each stratum. Since the sample mean is an asymptotically unbiased estimate of the population mean, range regression essentially models the trend of conditional expected value of the response as a function of ranges of explanatory variables. By controlling the strata, the method maintains the key source of variation and reduces confounding effects and survey bias intervening with the main explanatory variables. Using the FDA's HDS, we hypothesize that survey response bias may partially blur out the associations among BMI (body mass index, which involves body weight and body height) and food label use. Thus, we sort BMI into different ranges and then plot the mean responses across ranges to seek the relationship between BMI and consumer behaviors

on food label usage. After we apply the range regression technique, the associations between public perception on diet-nutrition and BMI range clearly stood out. Results of the new methods are compared with conventional approaches for model plausibility, goodness of fit, efficiency, and power performance.

Effects of Self-Awareness on Disclosure During Skype Survey Interviews **Shelley Feuer, *The New School for Social Research*; Michael Schober, *The New School for Social Research***

As people increasingly communicate via video using software like Skype and FaceTime, new opportunities for survey interviewing are emerging. But little is known about how videomediated interviewing affects data quality, respondent satisfaction, and interviewer rapport. On the one hand, videomediation might increase rapport with interviewers without the intimidation that can occur face to face; on the other hand, it may reduce respondents' sense of privacy, and thus reduce disclosure of socially undesirable behaviors. The current study explores how one prominent default feature in current video technologies—the “self-view,” a video image of oneself in the corner of the screen—affects survey respondents' levels of disclosure and feelings of comfort. In a laboratory experiment, 85 respondents engaged in a live real-time survey interview conducted over Skype, with the interviewer and respondent in separate locations. Respondents answered 42 questions from major U.S. surveys, selected because they might show mode effects related to socially undesirable responding, either with the default video image of themselves in the corner of the screen (“self-view”) or without the image (“no self-view”). Results suggest, perhaps counter intuitively, that the self-view reduces sensitivity and social desirability effects, allowing respondents to answer more comfortably and presumably more accurately. For instance, when asked about alcohol consumption, respondents in the self-view condition reported more frequent and greater alcohol consumption, and a (presumably more accurate) decreased likelihood of having been tested for HIV. In post-interview questions, respondents in the no-self-view condition reported a greater sense of co-presence with the interviewer and less comfort answering many of the sensitive questions. They also rated the interview as more sensitive than those in the self-view condition. Although the causal mechanisms are unclear, perhaps a self-view allows videomediated survey respondents to feel comfortable enough about their self-display to promote disclosure, or distracts them enough to reduce defensive self-monitoring.

Methodological Briefs: Maximizing Response and Response Quality

The Effect of Differential Mailing Methodologies on Response Rates: Testing Advanced Notices, Pre-Recorded Messages and Personalized Address Labels **Yelena Pens, *Arbitron*; Michelle Cantave, *Arbitron*; Robin Gentry, *Arbitron***

Arbitron Inc., a provider of radio ratings data, conducted a test using a probability based address sample to recruit the general population, aged 13 and older, to complete a one week Web-based diary of their radio listening. Since Web-based surveys historically have had lower response rates, there were several treatments in place to increase the response rate. In order to find the optimal mailing strategy for recruitment, the mailing experiment included treatments such as alternative advanced notices, pre-recorded telephone messages, and personalization. The initial invitation to participate in a one week Web-based diary included a box mailer with a monetary incentive. From previous testing, the box mailer provided the highest response rate as

compared to any reminder mailings. Thus, advance notices as well as pre-recorded messages emphasizing the arrival of the box mailer was the focus of the study. Three different package designs and messaging was tested which included a postcard, invitation note card, and self-mailer postcard. In addition, two different pre-recorded telephone messages were tested including an advanced notice message communicating the box mailer was on the way and a reminder message stating the box mailer had recently arrived in the mail. Finally, Arbitron previously conducted two studies testing how response rates and deliverability are affected by the use of generic salutation versus personalized name approach. The results of the studies were mixed, thus a follow-up study was conducted which included a name on the letters or a generic "City Name Area Household" greeting. In this presentation, we will present the results from the Web-diary initiatives. We will also determine the combined impact of the non-deliverable rate and response rate of the personalized letters. Finally, we will present the optimal mailing strategy for mail-based recruitment for an online survey.

New versus Old Technologies: An Examination of Usability and Cognitive Issues Across Modes Among Respondents with Varying Education Levels

Elizabeth M. Nichols, *U.S. Census Bureau*; Patricia L. Goerman, *U.S. Census Bureau*; Nathan Jurgenson, *U.S. Census Bureau*; Tiffany King, *RTI International*; Murrey Olmsted, *RTI International*; Jennifer H. Childs, *U.S. Census Bureau*

It has often been speculated that respondents who have lower levels of education may have trouble completing automated government forms. However, recent data shows that cell and smartphone usage is growing in this demographic (Woelfer, et al., 2011; Rice, et al., 2011; Woelfer and Hendry, 2009). With cell phone usage, in particular smartphones, becoming nearly ubiquitous, particularly among young people and minorities, there is the potential to use this technology to reach out to those with low education, who are often highly mobile and might, otherwise, not be included. However, little is known about the success and problems encountered in attempting to administer government forms via smartphones and tablets—in particular with those who are of differing educational levels. This paper presents qualitative evidence from 160 cognitive interviews completed with individuals who completed paper or automated versions of draft U.S. 2020 Census forms. The paper examines whether there are differences in the number and types of usability and cognitive problems found in cognitive interviews by education for paper and automated forms and seeks to identify whether data collection using automated mobile forms would be helpful in reaching out to those who have lower levels of education.

Converting Nonrespondents to Late Respondents: The Impact of Automated Phone Reminder in an RDD Landline Survey

Robin Gentry, *Arbitron*; Vrinda Nair, *Arbitron*

The Arbitron Syndicated Radio Survey uses a two-stage methodology whereby an RDD sample is contacted via telephone and all household members aged twelve or older are asked to participate in a seven-day radio listening diary for a specific "ratings" week. Unfortunately, roughly 40% of households who agree to participate in the Radio Survey during the phone call fail to return any diaries. Relatively little is known about why these households do not return their surveys. In Spring 2012, Arbitron fielded a study in which non-returning households were sent an automated phone message approximately 9 days after the end of their diary keeping week which reminded them to return their completed Radio Survey. We will present the return

rate results, cost-benefit analysis, as well as the analysis of the demographics of those that returned a diary to determine who we brought in with the additional automated phone reminder. Instances when the late respondents picked up their phones to receive the live automated phone message were compared to when the automated message was left on a voice mail to determine if there was a difference in sample performance.

Factors Influencing Survey Participation Rates on an Online, Probability-Based Research Panel

Dawn Wiest, *American College of Physicians*

In May 2011, the American College of Physicians (ACP), a membership organization of physicians who specialize in internal medicine, established a probability-based, invitation-only research panel to learn more about the needs and interests of members. After three waves of invitations, 952 ACP members had joined. In summer 2012, a process of “panel hygiene” was initiated with the goal of clearing the panel of non-participants and replacing them with a new round of invitees. Analysis revealed that 30% of panelists had completed no surveys or only one since joining. Brief surveys were sent to these panelists asking if they wished to remain on the panel. Panelists who did not respond to this survey and those who responded “no” were dropped from the panel. Beginning in October 2012, invitations to join the panel were sent to a new sample of ACP members. This five-minute presentation is based on an analysis of one year of panel participation data and highlights findings regarding participation rates and panelist retention. Over the course of one year, seventeen surveys were sent to panelists. Participation rates were influenced less by demographic factors, such as age, gender, or career stage, than by how soon after joining the panel panelists received their first survey. Forty percent of panelists who received their first survey over two weeks after joining completed no surveys in a year, compared to fourteen percent of those who received their first survey within ten days. The findings underscore the importance of minimizing the time between when a panelist joins a panel and when s/he receives the first survey. Additionally, analysis reveals that as a mechanism for engaging panelists, “quick polls” and other low incentive opportunities are no replacement for surveys offering higher value rewards. Recommendations based on the findings are discussed.

When We Do Not Know the Difference – the Level of DK in Different Question Formats and Different Modes

Steve Schwarzer, *TNS Opinion*; Eva Zeglovits, *University of Vienna*; Dylan S. Connor, *University of California (UCLA)*

The level of don't know (DK) responses recorded in surveys are impacted by both social desirability (SD) and satisficing (SC). Both, SD and SC are known to be sensitive to survey mode and can inflate the rate of non-committal responses. It is assumed that Web surveys mitigate interviewer effects, and thus social desirability. However, this is a dualism as Web surveys also tend to exhibit higher levels of Don't Know. This mechanic of survey design is poorly understood and there is little available, practical guidance on reducing mode effects that tend to increase the level of Don't Know selection. Our first research question addresses the level of don't know responses in Web surveys. We investigate how different presentations of don't know answers in this mode affect the number of respondents selecting those options. As many studies are now employed in a multi-mode manner, inconsistency in don't knows between modes introduce noise into the data. As such, our second objective takes a comparative

approach to modes, by analyzing the different outcomes between online and telephone surveys. To answer these questions we deployed a survey experiment, administered online in four countries (n=1000). So far, most studies have used data from lagged surveys. But in our case, the telephone benchmark surveys were conducted (n=1000) concurrently. The paper will focus on examining whether different question designs result in different outcomes in the level of don't know within the same mode. Furthermore, we will show, which question formats limit the differences between modes—online and telephone surveying. Finally, as this research is based on a multi-country survey, we will test, whether different formats work differently across countries. The paper will conclude with how researchers can successfully bridge modes in order to limit the “questionnaire design mode effect” on the answering behavior of respondents.

Data Quality in a Multi-Mode Self-Administered Study of Mental Health

**Andrew L. Hupp, *University of Michigan*; Margaret L. Hudson, *University of Michigan*;
Heather M. Schroeder, *University of Michigan***

This study examines important dimensions of data quality from a mental health study of soldiers in the U.S. Army. One component of this study involves a cross-sectional survey in which a global, representative sample of active duty soldiers is interviewed. Soldiers completed either a computerized or paper self-administered interview in a group session dependent on their duty location. Each group session is overseen by staff trained by an academic research organization. We will examine data quality using the following metrics: unit non-response (consent) rates, item non-response rates, a measure of satisficing (straight-lining) in responses to grid formatted questions, rates of endorsement of sensitive items, and questionnaire completion rates. This paper focuses on two aspects of the survey that may affect these measures of data quality. The first aspect examined is the impact of mode of administration. We hypothesize that data quality is improved when the survey is self-administered via computer rather than paper. The second aspect examined is the effect of the group administration agent (field staff v. Army). We hypothesize that data quality is affected by the presence of a homophilistic agent. In this case, the homophilistic characteristic of interest is being a member of the military. The agent is dressed similarly (Army uniform) to the participants. The agent may be perceived as an authoritative figure since they may have a higher rank than some of those being asked to participate. This could have an effect on perceived privacy and confidentiality by the participant, leading to higher compliance in completing the survey request while at the same time contributing to lower data quality through higher item nonresponse, more satisficing and less endorsement of sensitive mental health items.

Using Registry Information to Adjust for Non-response Bias in a Diabetes Patient Survey

Jiaquan Fan, *Mayo Clinic*

Objective: To evaluate nonresponse bias in a mail survey of diabetes patients and assess a weighting method designed to adjust for the non-response bias using information obtained from diabetes registry. Study Design and Setting: Patients from a diabetes registry including 34 Midwestern clinics were randomly selected to participate in a mail survey. 2055 patients responded to the survey (response rate 43%). Analyses examined demographic, current smoking status, and health outcomes: blood pressure, Hba1c, Low-density lipoprotein (LDL) from the diabetes registry, seeking differences between responders and non-responders. A logistic regression model is developed to identify significant factors related to non-response, and

a weighting method is designed to adjust for non-response bias. Results: Non-response bias is present in the survey. Responders tends to be older, nonsmoker, and healthier. Age, current smoking status, blood pressure, and LDL are identified to be significantly related to non-response. After imputation for missing values, these four variables were used to form weighting cells to create weights for non-response adjustment. This method compares favorably than non-response adjustment weighting using only demographic variables. Conclusions: Leverage salient theory suggests that topic is a large motivator for response. In practice few studies have frame data with which to conduct nonresponse bias analyses and weighting adjustments. When frames do have information on both respondents and non-respondents, it is typically only demographic variables and it is not clear how well adjustments made on demographic variables actually correct for observed bias in health related survey. Using rich information obtained in the registry database for this survey, we demonstrate that non-response in health related survey is likely related to health outcomes and registry data with rich health related information can be used to obtain better non-response adjustment than using demographic variables alone.

Issues Related to Recruiting and Screening

Empirical Assessment of Respondent Driven Sampling

Zeynep T. Suzer-Gurtekin, *University of Michigan*; Sunghee Lee, *University of Michigan*; James Wagner, *University of Michigan*

Challenges of scientific data collection with rare and hidden populations are well understood. Sampling such groups using traditional probability methods is highly costly and almost impractical. In order to approach this sampling issue, several methods utilizing the social networks of those populations, including respondent driven sampling (RDS), were suggested as an alternative. RDS stems from the reasonable assumption that, although hidden in the general population from outsiders' viewpoint, some hidden population units are linked to other units of the same population, forming some type of networks. Once a few members of the target rare population are contacted typically through convenience sampling, those members are interviewed as first-wave participants (seeds) and their social networks are exploited to recruit the next wave of the participants. Unlike traditional sampling, these seeds are asked to play a role of recruiters; they recruit those who qualify for the study from their individual networks. After the second wave of data collection, this new set of participants recruits the next wave of participants. Recruitment waves continue until the desired sample size is achieved. Under a set of strong, yet often untestable, assumptions, RDS claims to produce memoryless Markov chains of data points leading to unbiased estimates. In this paper, we use data from the Sexual Acquisition and Transmission of HIV Cooperative Agreement (SATHCAP) that collected data from the HIV risk groups using RDS. We examine how well the assumptions are reflected in the data collection, focusing on the memoryless chain assumption and the complete response assumption. The examination is done with respect to estimation and sampling productivity. We also compare different estimators suggested in the literature to test their performance.

Recruiting Participants into a Probability-Based Panel Using Interactive Voice Response Methods: The Canadian Experience

Frank L. Graves, *EKOS Research Associates*; Timothy B. Gravelle, *PriceMetrix Inc.*

Significant research on recruiting participants into probability-based research panels has been undertaken in recent years. In particular, research has focused on finding optimal recruiting

processes and assessing the representativeness of samples recruited using different methods -- landline random-digit dial (RDD), dual-frame (landline RDD plus cell phone) and address-based sampling (ABS). To date, little work has been done to evaluate the relative efficacy of interactive voice response (IVR) methods, in part due to regulations in the United States preventing IVR dialers from calling cell phones and the bias that would presumably result from using IVR methods to call landline RDD sample only. This paper presents experiences and findings from the use of IVR to recruit into a probability-based panel in Canada, where both landline and cell phone numbers may be called using IVR.

Benefits and Drawbacks of a Multistage Screening Effort for Surveying Rare Populations

Heather M. Morrison, *NORC at the University of Chicago*; Alicia M. Frasier, *NORC at the University of Chicago*; Stephen J. Blumberg, *National Center for Health Statistics*; Matthew D. Bramlett, *National Center for Health Statistics*

Conducting scientifically rigorous surveys of rare populations can be cost-prohibitive because obtaining a sufficient sample of eligible respondents via probability sampling requires a significant screening effort. As a result, surveys of rare populations are sometimes undertaken using convenience samples that minimize the screening effort but come at the cost of scientific rigor. Recent survey work undertaken through the State and Local Integrated Telephone Survey (SLAITS) mechanism of the National Center for Health Statistics, however, demonstrates it is possible to control screening costs while maintaining the statistical properties of a probability design. SLAITS' multi-stage approach screens for rare populations via one or more parent surveys: the National Survey of Children with Special Health Care Needs and the National Survey of Children's Health – both conducted on behalf of the Maternal and Child Health Bureau. These national surveys use the National Immunization Survey sampling frame to screen approximately six million telephone lines for eligible households yearly, resulting in a rich sample of certain rare populations. Once identified, these targeted rare populations participate in the salient follow-up survey. We have successfully employed this screening methodology to identify and interview nationally representative samples of adoptive parents in the National Survey of Adoptive Parents and the National Survey of Adoptive Parents of Children with Special Health Care Needs and more recently for parents of children with autism, intellectual disability, or developmental delay in the Survey of Pathways to Diagnosis and Services. These surveys would not be feasible without this multi-stage screening mechanism. There are, however, drawbacks to this approach. While observed cooperation rates are high for the salient survey, response rates must be calculated accounting for response at all survey stages including screening. We examine the benefits and drawbacks of interviewing rare populations using this methodology, including assessing survey cost, response rates, and sampling alternatives.

Assessing Methods of Recruitment for a Cell Phone Survey Panel: An Experiment Conducted in 2011 in Mexico City

Yamil Nares, *University of Essex*; Rene Bautista, *NORC at the University of Chicago*

This paper presents the results of an experiment conducted with cell phones in Mexico City between July and August of 2011. The study was conducted by the public opinion firm Defoe, Experts on Social Reporting and consisted of a three-wave survey. In the first wave, a household survey of one hundred cases was conducted face to face, as a baseline study.

These selected respondents were provided with free pre-paid cell phones in exchange for their continued participation in subsequent waves, which were planned to be conducted over the said cell phones during the following week. The pool of selected respondents was randomly divided into two groups. Fifty of the respondents were handed a letter which states the purpose of the study and objectives. The other fifty people were asked to voluntarily sign a contract in order to encourage commitment and participation over the next two waves. In both conditions (letter and contract) cell phones were credited with 15 dollars in advanced. Participants were explained that they could keep the cell phone equipment upon completion of the two-wave study; that is, by the end of the week. This paper will discuss the impact of using signed contracts (compared to letters only) on survey participation. Aspects such as interviewer characteristics, fieldwork data, and other relevant information will be included in the analysis.

Strategies for Recruiting Respondents for Exploratory Interviews to Aid Questionnaire Development

Herman Alvarado, *U.S. Census Bureau*

In a recent collaborative effort between the National Science Foundation and the U.S. Census Bureau, a sample of U.S. companies was contacted to understand the role of innovation in their business practices and decision-making, to assess the feasibility of developing survey questions to measure private-sector innovation. This type of exploratory research is often foreign to potential research participants, and may even be viewed with suspicion. Thoughtful, concise and persuasive appeals are often necessary to find, contact, and obtain cooperation from appropriate people within companies. In order to interview appropriate company personnel, i.e., those with both broad and deep knowledge of their companies, we decided to make the initial requests to company executives and ask for their assistance. In order to reach company executives, an initial mail contact strategy was used. An official letter explaining the purpose of the study, requesting their participation, and providing the researchers' contact information was sent to more than 120 companies in several U.S. cities. We took steps to ensure the letters would be perceived as legitimate and important, and would get the attention of gatekeepers responsible for filtering executives' mail, including personalizing the letters and sending them via 2-day priority mail. We conducted telephone follow-ups with those companies who did not initially respond to the letters. In our presentation we will discuss recruiting strategies and methods, as contacts with often busy and skeptical company representatives, especially gatekeepers, present narrow windows of opportunity to convey the nature of the request for the interview. We will also make recommendations for overcoming some of the obstacles we encountered.

Multi-Mode Surveys

Evaluation of a Sequential Mixed-Mode Design Experiment with Physicians on Response Rates, Costs, and Response Bias

Emily Geisen, *RTI International*; Murrey Olmsted, *RTI International*; Joe Murphy, *RTI International*; Marshica Stanley, *RTI International*

While Web surveys are generally less expensive than data collection by mail, they have not been shown to be successful at achieving high response rates with physicians. In comparisons of single-mode physician surveys, Web surveys typically have lower response rates than other modes (Van Geest, 2007). Similarly, research on concurrent mixed-mode surveys with

physicians has found that the use of a Web option does not increase survey responses compared to mail alone (McFarlane, 2009). However, a recent meta-analysis of mixed-mode general population surveys found that offering sequential mixed modes (offering only one mode at a time) compared to concurrent mixed modes (offering only more than one mode at the same time) can yield higher responses rates (Fulton, 2012). Our study evaluated a sequential mixed-mode design experiment conducted on a nationally representative sample of 4,700 board-certified physicians. Recent research with physicians shows that physicians are adopting mobile devices such as smartphones and tablets at increasing rates. Therefore the Web survey was optimized so that it could be completed on mobile devices as well as computers. Half of the sample received an initial paper survey via mail followed by up to three mail-only nonresponse follow-ups. The other half of the sample received an initial survey invitation via email with up to two email reminders. Nonresponders to the Web survey were then sent up to three paper survey follow-ups. The three paper survey follow-ups were identical in both groups. In this paper, we compare the effect of the two mixed-mode designs on responses rates, overall costs, and costs per complete. In addition, we examined mode differences and potential effects of response bias between the two groups. This work has implications for researchers designing studies with physicians to find an optimum balance between costs and response rates.

Facing Their Fears: Examining the Impact of Audio Computer-Assisted Self Interviewing on Population Prevalence of Self-Reported Non-Specific Psychological Distress

Sarah S. Joestl, *National Center for Health Statistics*; James Dahlhamer, *National Center for Health Statistics*; Adena Galinsky, *National Center for Health Statistics*; Marcie Cynamon, *National Center for Health Statistics*; Virginia Cain, *National Center for Health Statistics*; Jennifer Madans, *National Center for Health Statistics*

Despite steady growth in psychiatric epidemiological research, population-based prevalence estimates of serious mental illness remain the gold standard for both research and policy. Recognizing this need, the National Center for Health Statistics (NCHS) during its redesign in 1997 added the K6 scale, a validated six-item screening tool for identifying non-specific psychological distress, to its National Health Interview Survey (NHIS). However, concerns around stigma and discrimination may disincentivize people living with mental illness from reporting psychiatric symptoms in a face-to-face or telephone interview setting. In order to assess the possibility of underreporting (and hence bias of national estimates) of this and other health information, NCHS between August and mid-October 2012 carried out a feasibility study on the use of Audio Computer-Assisted Self Interview (ACASI) for a subset of NHIS questions deemed sensitive in nature. In this paper, we used data from that field test to compare prevalence, item non-response, and breakoff rates for each of the K6 items and the overall scale between the 3,215 adults who received the questions via ACASI and the 2,237 adults who completed them via Computer-Assisted Personal Interview (CAPI). We further contrasted CAPI field test estimates with those from the 2012 NHIS production survey to allow examination of potential context effects from changes in item placement within the survey. Where significant bivariate results emerged, we examined them in multivariate models to identify potential sociodemographic respondent characteristics underlying any observed mode effects. Results from this examination will not only inform mode choice for future surveys with a mental health component, but will also provide insight on whether prior-year NHIS estimates of non-specific psychological distress could be improved to account for context effects due to question placement.

Alone in a Group: Comparison of Effects of a Group-Administered Paper-Pencil Survey Versus an Individually-Administered Web-Based Survey on Perceptions of Culture, Peer Pressures and Stigma

William B. Higgins, *ICF International*; Frances M. Barlas, *ICF International*; Jacqueline Pflieger, *ICF International*; Randall K. Thomas, *GfK Custom Research North America*; Diana Jeffery, *Tricare Management Activity*; Mark J. Mattiko, *United States Coast Guard*

While research has found that the presence of an interviewer can influence respondents' answers to questions, less attention has focused on the potential impact that other respondents may have on survey responses as might occur in group-administered settings. In assessing topics related to group culture and peer-pressure, the presence or absence of other group members when completing the survey may influence responses. Such influences may be stronger in a tight-knit group like the United States military where unit cohesion and trust are critical to mission success. In this study, survey responses to items concerning group culture and influence when asked on a paper-pencil, group-administered survey were compared with responses on an individually-administered, online survey. The Department of Defense and U.S. Coast Guard authorized the 2011 Health Related Behaviors Survey to explore the prevalence of a number of behavioral health issues including the military culture of substance use, the presence of peer pressure to use substances, and the stigma associated with receiving mental health services. Personnel from a few key military installations from the Army, Navy, Marines, and the Coast Guard were randomly assigned to one of the administration modes. Respondents were assured anonymity for each mode. Group-administered paper-pencil survey respondents indicated greater stigma of receiving mental health care and a stronger military culture of substance use than did respondents in the Web-based mode.

The Effect of Survey Mode on Socially Undesirable Responses to Open Ended Questions: Online vs. Paper Instruments

Eric Hedberg, *NORC at the University of Chicago*; Gabriel Ceasar, *Arizona State University*; Danielle Wallace, *Arizona State University*

A chief concern of survey research is that respondents give socially desirable answers instead of actual beliefs. However, it is possible that this tendency is mitigated by survey mode. In this paper we evaluate open-ended responses to a photographic stimuli that asked 1,056 students in a criminal justice program to evaluate neighborhood conditions. This photograph presents a street corner with a brick building, a van marked with spray paint, and a religious mural. We expect responses to this photograph to contain references to race, ethnicity, and class. However, we examine the difference in how race, ethnicity, and class, were depicted by respondents across two modes: paper surveys (46.6 percent of responses) and Web surveys (53.3 percent). We mark each response for various socially undesirable responses ranging from impolite language to disparaging stereotypes. We then use an item response theory (IRT) model to estimate the impact of survey mode on the propensity of such offences by estimating a multi-level logistic regression model. Using a means-as-outcome model and cross level interactions with survey mode we estimate how mode impacts not only the general propensity for social undesirability, but also how survey mode impacts the different aspects of socially undesirable answers. Preliminary results suggest that while mentions of race or ethnicity do not vary based on mode, surveys from Web interfaces are more likely to provide socially undesirable answers. For example, we found no difference between modes for mentions of minority populations, but online surveys were 88 percent more likely to use the word "ghetto."

We then consider what these results suggest for quantitative research. We conclude that online surveys are more likely to elicit visceral responses, and that analyses on mixed mode data collection should include survey mode as a control when examining mean differences on various scales.

Mode Effects in a National Establishment Survey

Kelly Daley, *Abt SRBI*; Ben Phillips, *Abt SRBI*

Surveys of establishments often require the reporting of administrative or historical data, which can be difficult or burdensome to complete by telephone. Offering survey respondents multiple modes of reporting can make the task easier by allowing respondents flexibility in the time, location and pace at which they complete the survey. Presumably, this flexibility would increase response rates, produce higher quality data and potentially reduce survey administrative cost. The 2012 Family and Medical Leave Worksite Survey was a sequential multi-mode (Web and CATI) survey of 1,812 U.S. business establishments. A major design difference between the 2012 survey and earlier administrations is that the 2012 survey allowed respondents to complete the survey on the Web. The field period for the 2012 survey was March through June, 2012. A total of 634 interviews were completed on the Web and 1,178 interviews were completed by Computer Assisted Telephone Interviewing (CATI). The target population consisted of all private-sector business establishments excluding self-employed businesses without employees, government entities, and quasi-government entities. Provision of the Web option in 2012 was expected to bolster both the overall response rate and the item response rate on several key variables related to the administration of FMLA at the sampled establishment site. This paper explores several aspects related to survey administration mode in the 2012 FMLA Worksite survey. We compare item response rates to administrative data questions between the 2000 and 2012 surveys. We examine mode effects using matching models for causal effects due to the non-ignorable relationship between respondent characteristics and completion of a survey in telephone or self-administered modes. Potential reduction of bias of estimates due to different sample composition under a high response rate scenario is estimated net of estimated mode effects.

Applications of Social Media to Surveys and Pretesting

Social Media vs. Online Classified Advertisements: Does Where We Advertise for Cognitive Interviews Matter?

Brian Head, *RTI International*; Elizabeth Dean, *RTI International*; Timothy Flanigan, *RTI International*; Jodi E. Swicegood, *RTI International*; Michael Keating, *RTI International*

Technologies have advanced over the past decade and the ways in which people access information has evolved with those advancements. These changes have created new opportunities to recruit questionnaire evaluation study participants (e.g., cognitive interview participants) that may address some concerns with the use of one of the most common recruitment methods in use today—online classified ads (e.g., advertisements on Craigslist.com). Potential issues with online classified ads include: the recent decline, based on anecdotal observation, in the number of responses to these ads; limited demographic diversity; an inability to target specific populations; concerns about the development of a class of “professional participants” who use the ads to seek out study participation for additional income; and the infeasibility to recruit geographically dispersed samples. We hypothesize that

advertising on social media may help address these concerns. We use recruitment data from two cognitive interviewing studies with distinct populations—virtual world users and adults near retirement age—to test this theory. In both studies we ran advertisements on Facebook and Craigslist to recruit potential study participants. Each ad included a link to a web administered screening survey. The screening surveys included questions about demographic information and other information used to determine study eligibility. We will present data showing differences in 1) demographic diversity of participant pools drawn from the two recruitment methods; 2) the size of and speed at which pools are drawn; and 3) the feasibility of recruiting a geographically diverse population. Findings from this study may be useful to researchers concerned 1) with the effects of having homogeneous pools from which to draw questionnaire evaluation participants; 2) the effect professional participants may have on cognitive interviewing data; and, 3) recruiting geographically dispersed pool of potential study participants.

Cognitive Interviewing in Online Modes: a Comparison of Data Collected in Second Life and Skype

Jodi E. Swicegood, *RTI International*; Brian Head, *RTI International*; Elizabeth Dean, *RTI International*; Michael Keating, *RTI International*

Cognitive interviewing can identify potential errors in a survey prior to a large data collection effort allowing researchers to effectively pretest a draft survey instrument. Digital technologies afford researchers the opportunity to overcome geographic and logistical limitations of conducting these interviews with a diverse sample. The convenience of interviewing participants online includes reduced travel time and the ability to schedule interviews outside of normal business hours, reducing participant burden with certain populations including online users. The Second Life population was of interest to researchers in this study. Second Life is a virtual world where users self-represent through avatars. Purposes of play include socializing, entertainment and education. New technologies such as the virtual world Second Life and the voice-over-internet software Skype were utilized to conduct cognitive interviews pretesting a draft instrument on virtual world avatar similarity. A series of questions asked participants to describe several physical and personality characteristics of both themselves and their avatars. The goal of this questionnaire was to determine the extent to which SL users viewed their avatars as similar to their real life counterparts. Interviews were conducted in three modes: Second Life, Skype and face-to-face. To determine the feasibility of conducting cognitive interviews digitally, analyses were conducted to compare data quality across each mode; analyses identified the number, type and severity of errors detected. Preliminary findings suggest that interviews conducted in Skype and Second Life yield, on average, the same number of errors. Comparison data are presented from all three modes. Second Life and Skype can be used to conduct cognitive interviews with a sample of online participants, though each mode has its own consideration and limitations for study design and implementation. These implications are discussed and recommendations explored for researchers interested in other digital cognitive interviewing modes.

Latent Characteristic Extraction from Twitter Data: Toward Weighting Social Media Data to Make Inferences to the General Public **Martin Barron, *NORC at the University of Chicago***

Twitter is a social media service where users post small, 140-character public messages. In the U.S. alone, Twitter currently has over 100 million users who each day posts over 400 million “tweets”. This continuous stream of data has been mined by researchers to measure a variety of behaviors and opinions such as influenza outbreaks, drug use, and a host of other topics of interest to survey researchers and their clients. However, significant questions remain regarding the generalizability of these findings beyond the particular universe of Twitter users. Twitter users represent a self-selected cross-section of the U.S. population—a cross-section that is younger, more African American, and less rural than the overall U.S. population. One possible approach to drawing inferences from Twitter data to a larger population involves weighting the data drawn from Twitter users to known demographic distributions among the general population. Unfortunately, almost no demographic data are available on Twitter users. This paper describes a method of assigning demographic characteristics to Twitter users as a first step towards weighting data mined from Twitter to U.S. population control totals. I discuss a methodology for extracting latent characteristics (such as sex, race, and age) based on Twitter behavior. Starting with a hand-coded training dataset, I use machine learning techniques to build models classifying users on each demographic characteristic of interest. I show that, given a robust training dataset, many demographic characteristics can be assigned with relatively high levels of accuracy. Using these classification models, I then explore weighting election projections based on Twitter data to determine if the weighted results result in more accurate predictions than the unweighted projections.

Capabilities and Considerations for Using Facebook in Survey Research **Kim Mook, *Mathematica Policy Research*; Sean Harrington, *Mathematica Policy Research*; Amanda Skaff, *Mathematica Policy Research***

In an era of declining survey response rates and unreliable locating methods, social media provides important new opportunities for respondent outreach. With over 900 million users worldwide, including over 150 million in the United States, Facebook warrants particular attention as a tool for improving sample member contact. This paper discusses the potential capabilities and concerns that survey researchers must consider when exploring ways to incorporate this widely used platform into data collection and respondent locating efforts. We detail the demographics of Facebook’s most common users, as well as the benefits and drawbacks of contacting potential sample members on the site. We also describe Facebook’s current outreach capacities, including the differences in information dissemination and direct communication capabilities between a Page, profile and group, and the important privacy issues that circumscribe interaction on the site. Finally, we provide a brief case study of the preliminary stages of Facebook use on the Evaluation of the YouthBuild Program. We detail the benefits of using Facebook to locate and contact this study’s sample members, who are generally young, low-income, highly mobile, and often maintain social media accounts as their most permanent method of contact. In addition to these outreach strategies, we describe the development of tools to track social media interactions, as well as paradata possibilities for future exploration. Though these social media efforts are ongoing, our progress to date suggests that Facebook can be a critical tool in establishing connections with difficult to reach sample members, and can provide otherwise inaccessible contact information to locators in addition to serving as a communication platform.

Dangerous Disconnects? How Public Discourse About Nanotechnology is Missing the Point

Sara K. Yeo, *University of Wisconsin - Madison*; Dominique Brossard, *University of Wisconsin - Madison*; Dietram A. Scheufele, *University of Wisconsin – Madison*; Michael A. Xenos, *University of Wisconsin - Madison*

In general, scientists tend to be more optimistic about technologies, such as nuclear power and biotechnology, and perceive fewer risks than lay audiences (Savadori et al., 2004; Sjöberg, 1999). However, there is evidence that this trend is reversed for environment, health, and safety (EHS) risks of nanotechnology (Scheufele et al., 2007), with scientists calling attention to the potential seriousness of these negative effects. Although nanotechnologies are gaining in consumer uses with now over 1,300 products available worldwide (The Project on Emerging Technologies, 2011), the potential deleterious effects of nanoparticles on EHS have been gaining attention among scientists and regulators (Holgate, 2010; Marambio-Jones and Hoek, 2010). The extent to which these discussions have reached broader segments of the American population is an empirical question. In this study, we explore public discourse of nanotechnology using the micro-blogging social media platform Twitter. Online media are rapidly becoming an important source of information for science and technology for lay audiences (National Science Board, 2012). Twitter is one of the most prolific outlets for public discourse as it is an ideal medium for information distribution and discussion. For example, on the night of the 2012 U.S. Presidential election, 31 million tweets were posted, with the highest tweeting rate (327,452 tweets per minute) occurring when media networks announced Obama's reelection (Sharp, 2012). In the present study, we performed opinion mining with the software ForSight to characterize 1,557,325 nano-related tweets posted between September 1, 2010 and August 31, 2012. The topics analyzed included business, national security, consumer products, medicine/health, EHS, basic research, and energy, the most invested domains of research and development related to nanotechnology. We found that discussions about consumer products and national security dominate public discussions about nanotechnology, while EHS was the least discussed. Implications of this disconnect between expert and public discourses are discussed.