AAPOR
AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH

71st Annual Conference

2016 Conference Abstracts

Reshaping the Research Landscape: Public Opinion and Data Science

AAPOR 71st Annual Conference

May 12 – 15, 2016

Hilton Austin • Austin, Texas

www.aapor.org/conference [f] [in] [y] #AAPOR

Recruitment Challenges in Online Data Collection

Bees to Honey or Flies to Manure? How the Usual Subject Recruitment Exacerbates the Shortcomings of Non-probability Samples
Steven A. Snell, *Duke University*
D. Sunshine Hillygus, *Duke University*

Among the criticisms of online panels is their use of nonrandom recruitment and repeated interviewing of a subject pool. A growing literature has started to examine the data quality implications of so-called "professional" respondents—heavy survey takers thought to seek out surveys for the cash and incentives offered. While this literature has reached inconsistent conclusions about satisficing behavior among experienced survey takers, our previous examination of political outcomes by survey experience found professional respondents were less politically interested, engaged, and knowledgeable than other respondents. The observed patterns suggested countervailing biases associated with self-selection, with professional survey-takers motivated by incentives and nonprofessional survey-takers motivated by interest in the survey topic. The proposed experiment will build on this previous work by explicitly evaluating the impact of respondent's survey taking experience and motivation for participation on survey estimates. We employ a two-by-two experimental framework, in which we manipulate the description and advertised topic of the survey and the financial compensation offered to participants; implemented with two different online panels. We compare respondents across these four conditions in terms of their demographic profiles, political attitudes, and respondent quality, as well as their experience taking surveys and panel tenure. The proposed analysis will bridge and contribute to both the recent literature on professional respondents and the classic literature examining the consequences of interest in survey topic.

Exploring the Gaps in the Gig Economy Using a Web-based Survey: Measuring the 'And' that Captures Formal 'and' Informal Work Activity
Barbara J. Robles, *Federal Reserve Board*
Marysol McGee, *Federal Reserve Board*
Jenny Schuetz, *Federal Reserve Board*
Alexandra M. Brown, *Federal Reserve Board*

Many more American workers are involved with alternative, informal and contingent work arrangements than in previous decades (Schor, 2014; Friedman, 2014). Yet the usual employment data sources such as the Census Bureau's Current Population Survey (CPS) and the BLS' Current Employment Survey (CES) are limited in their ability to measure the growth of 'gig' and 'supplemental' work. Researchers and policymakers lack reliable data on some basic questions about informal work, such as: how many Americans engage in occasional work? For those who do occasional work, how many 'gigs' do they have, and how much income do they derive from gigs versus traditional labor income? What are the demographics of those most involved in alternative work arrangements? Is the increase in occasional work driven by "push" factors (economic necessity and financial insecurity) or "pull" factors (preference for flexibility)? We provide preliminary evidence on these and other questions related to

informal work, employing data collected from the GfK General Population Panel, a national internet panel of approximately 2,000 adult respondents with an oversample of low-to-moderate income households (ie., zero to $40K). Respondents are asked a detailed series of questions about their work modes, including whether they work full time 'and' part time, whether they work for a traditional employer 'and' for themselves, and whether they undertake occasional 'gig' work. We measure respondents' assessments of the significance of their 'gig' work on supplementing household monthly budgets. The nature of the occasional 'gig' work covers online oriented tasks (selling services) and using the web to sell new or used goods (eg. eBay, Etsy, etc.), ridesharing (eg., Uber, Lyft, etc.) and renting space or items (eg., AirBnB, etc.). We also capture the more traditional venues for supplementing income: flea markets, thrift and consignment stores, mobile food trucks and carts, etc.

The Dynamic Composition of Amazon Mechanical Turk Samples

Jesse J. Chandler, *Mathematica Policy Research*
Logan Casey, *University of Michigan*
Adam Levine, *Cornell University*
Dara Strolovitch, *Princeton University*
Andrew Proctor, *Princeton University*

Amazon Mechanical Turk (MTurk) has generated considerable interest as a low-cost platform for survey and experimental research. A well-known limitation of MTurk is that its population of workers is not representative of the US population. It is often overlooked that individual samples of workers are also not representative of the population of MTurk workers as a whole. To examine potential differences across MTurk samples, a large (N ~10,000) sample of workers was recruited through postings on different days of the week and at different times of day (10am, 3pm and 10pm) over a period of two months. Workers reported demographic information about themselves and completed a brief personality inventory. Workers' prior experience completing research studies was estimated using a database of ~100,000 previously recorded responses. Worker characteristics were examined as a function of day, time of day and serial position in which respondents completed the survey. Differences in time zone, mobile device use, and demographic characteristics emerged as a function of the day and time of day that data were collected. Differences also emerged as a function of serial position, demonstrating that sample composition systematically changed as sample size increased. These differences may be of particular importance to researchers interested in MTurk workers as a population in its own right (e.g., organizational researchers interested in the "gig economy") and those seeking to recruit convenience samples of specific compositions. Substantial differences in worker's prior experience with research emerged across day, time of day, and serial position of responses, reflecting that "professional" respondents complete surveys sooner and are particularly active earlier in the day and in the latter half of the week. This finding is of more general importance to those who wish to limit the proportion of professional respondents in a sample.

Recruitment of Targeted Populations via Social Media: Examination of Non-probability Based Sampling Approaches

Ipek Bilgen, *NORC at the University of Chicago*
Ilana Ventura, *NORC at the University of Chicago*
Michael J. Stern, *NORC at the University of Chicago*
Angela Fontes, *NORC at the University of Chicago*

Declining response rates and increasing data collection costs are adversely affecting probability-based surveys. Hence, online-based sampling and recruitment methodologies have emerged and have been examined as an alternative approach to address cost, non-response, and coverage concerns related to the probability-based sampling strategies. The dramatic national and global increase in social media use and its potential to lower cost per complete allow this platform to be considered as a potential sample recruitment venue. Previous research has shown that social media sites can be used for recruitment to web surveys yet the findings are equivocal and suffer from two shortcomings. First, they tend to rely on only one social media site (e.g., Facebook). Second, they rarely examine the usefulness of social media platforms for their ability to reach targeted populations. Our paper reports findings from two studies that used social media and search engine ad campaigns to recruit targeted populations for web-based surveys. In our first study, we employed targeted ad campaigns via Google, Facebook, and Twitter to recruit 18-25 year olds to complete a web survey regarding health and social life. Given the success of Facebook reaching targeted populations in our first study, our second study focused on recruiting another targeted population (Asian-Americans) to a web survey on financial well-being via Facebook-based ad campaigns. We examine differences in response rates by platform, as well as the substantive responses of our respondents in comparison to similar questions asked in general population probability based surveys. We further explore how each of the social media and search engine platforms perform relative to each other. Ultimately, the results expand our understanding of using social media and search engine ads for targeted survey recruitment both in terms of their strengths and shortcomings.

Writing and Formatting Questions to Improve Data Quality

The Effects of Respondent and Question Characteristics on Respondent Behaviors

Amanda Ganshert, *University of Nebraska-Lincoln*
Kristen Olson, *University of Nebraska-Lincoln*
Jolene D. Smyth, *University of Nebraska-Lincoln*

In standardized telephone interviews, the interviewer ideally reads the question, including response categories, as written and without error, followed by the respondent providing an answer that fits the response task (Fowler & Mangione 1990; Blair & Srinath 2008). However, there are many deviations from this ideal, including respondents providing inadequate or other types of answers, requesting clarification, or otherwise talking with the interviewer. Some questions (e.g., cognitively burdensome, lengthy questions) may be more prone to these respondent behaviors than others. Generally, evaluations of question wording look at one characteristic at a time; yet questions are comprised of multiple characteristics, some of which are easy to experimentally manipulate (e.g., number of response options) and some of which are not (e.g., attitude vs. behavior). All of these characteristics can affect how respondents answer these questions. Using a landline telephone interview conducted by AbtSRBI during the summer of 2013 (n=450; AAPOR RR3=6.3%), we use cross-

classified random effects models to simultaneously evaluate the effect of multiple question, respondent and interviewer characteristics on respondent behaviors. We look at question characteristics (long questions, questions with high reading levels, etc.), respondent characteristics (e.g., education level, age), and interviewer characteristics (e.g., experience). Interviewer and respondent behaviors were coded at the conversational turn level, and then aggregated to the question level for this analysis. Preliminary analyses show that "don't know" responses are higher for questions with higher reading levels, questions earlier in the questionnaire, and attitude questions. Inadequate responses are more likely to happen on behavioral questions and for earlier questions in the questionnaire. In addition, respondents with high school education or less answered inadequately or "don't know" more often than those with more education, as did older people. This paper will conclude with the implications of these findings for questionnaire design.

Mismatches
Jolene D. Smyth, *University of Nebraska-Lincoln*
Kristen Olson, *University of Nebraska-Lincoln*

Many questionnaire design texts emphasize that the wording of survey questions needs to (1) express the concept being measured and (2) tell respondents how to answer the question adequately (e.g., Dillman, Smyth, & Christian, 2014; Fowler, 1995). Researchers tend to focus heavily on question wording but sometimes overlook the response task itself, resulting in question wording that does not match the response options provided (i.e., mismatches). Three common examples are yes/no questions with ordinal or nominal response options (e.g., Do you agree with topic x? Strongly agree, agree, etc.), open-ended questions with closed-ended response options, and check-all-that apply questions with forced-choice response options. Such mismatches likely affect data quality, but may have different effects in self and interviewer-administered modes because respondents can see the response options in visual, but not auditory, modes. In this paper, we report the results of mail and telephone experiments in two surveys - the National Health, Wellbeing, and Perspectives Study, a mail survey conducted during the summer 2015 (n=6,000 sampled; n=1002 responded; AAPOR RR1=17.8%) and the Work and Leisure Today II survey, a telephone survey also conducted during the summer 2015 (n=911, AAPOR RR3=7.8%). We examine response distributions and data quality indicators (e.g., item nonresponse, straightlining, and marking only in the "yes" column in a forced-choice item) in questions where the question stem and response options are matched versus mismatched. Preliminary analyses indicate that the questions in which the question stem does not match the response task conveyed in the response options resulted in significant and substantively meaningful differences in the distributions of survey responses. These mismatches also tripled the item nonresponse rates compared to the questions where the response options were aligned with the stem. The paper will discuss implications for questionnaire design in both mail and telephone surveys.

What Happens When Respondents Don't Know the Answer to a Survey Question? (Hint: They Do Not Answer 'Don't Know.')
Stephanie Willson, *National Center for Health Statistics*

Respondents do not always know the answer to survey questions posed to them. However, many survey administration protocols suppress the 'don't know' option in favor of encouraging respondents to choose an answer from among the response categories. This paper discusses how respondents who do not know the answer to a survey question decide to respond when 'don't know' is not an apparent response option. Findings are derived from a study done by the Collaborating Center

for Questionnaire Design and Evaluation Research, National Center for Health Statistics on survey items aimed at measuring the prevalence of cervical and lung cancer screening procedures. The study was a cognitive interview evaluation of cervical cancer screening (Pap and HPV tests) and lung cancer screening questions (x-rays and CT scans). Forty interviews were conducted. The preeminent finding was that many respondents had little-to-no knowledge of the tests they had received. They did not always know or remember the names of medical tests they had, nor did they always know or remember the purpose of the tests. When respondents answered questions about cervical and lung cancer screening procedures, their lack of knowledge had a bearing on the judgment stage of the question-response process. Absent the ability to report 'don't know,' they employed one of two judgment strategies. On strategy was to base their answers on assumptions or speculations of what might be true. For example, an answer of 'yes' was given when respondents had multiple tests. A second strategy was to choose the response category that best served as a proxy for 'don't know.' For example, an answer of 'no' was often given when respondents thought they would have remembered if they had the test. These patterns of response -- and their potential effect on survey estimates -- will be discussed in more detail.

The Matrix Revolution: Improving Grids for Online and Mobile Surveys
Randall K. Thomas, *GfK Custom Research*
Frances M. Barlas, *GfK Custom Research*
Nicole R. Buttermore, *GfK Custom Research*

In self-administered surveys, grids have evolved to efficiently measure multiple concepts using the same graded response format across a number of different targets, with responses typically in columns and the rating targets to evaluate in rows. Generally, grids allow us to more efficiently assess multiple targets compared to presenting each target is presented on separate screens. The response format that is chosen can vary in terms of semantic labeling - every response can have an associated semantic label (fully anchored) or only the end categories have a semantic label (end anchored). The response format could represent a bipolar scale (ranging from a concept and its antithesis, e.g. like – dislike) or a bipolar scale (with a concept and its absence, e.g. like – do not like). Finally, the response format can vary in terms of the number of response categories used. First, we summarize 4 studies comparing the use of end-anchored and fully-anchored categories with both unipolar and bipolar scales, holding the number of categories constant. The unipolar fully labeled variant generally appears to be better as a measure across topics, with better differentiation and validity. Since online surveys are increasingly being taken on mobile devices, the utility of a grid is becoming affected by screen size. In a series of 2 additional studies we examined how reducing the number of grid response categories affects the measurement of a variety of concepts. We found that response formats with fewer response categories take less time to complete, are easier to complete on mobile platforms, and can show as much validity as formats with more response categories, especially when using the unipolar variant. We examine the ability of the scales to detect smaller differences between rating targets and found some improved differentiation with shorter scales rather than longer scales.

New Ideas in Using GIS & Remote Sensing Data for Survey Research

New Ideas in Using GIS & Remote Sensing Data for Survey Research
Stephanie Eckman, *RTI International*

The application of Geographic Information Systems (GIS) and remote sensing data to survey data collection and analysis has dramatically increased in recent years. As these technologies become less expensive and easier to use, and geographic data becomes more widely available on the web, we expect survey researchers to find even more uses for these tools. While we embrace these tools, however, we should also maintain a healthy skepticism about their capabilities and limitations. This session brings together innovative work by researchers at diverse organizations who are thinking creatively about how to use these technologies to collect survey data, analyze survey data, or make surveys more accurate and efficient. The papers represent the breadth of the applications of GIS technology to survey research.

Spatial Modeling through GIS to Reveal Error Potential in Survey Data: Where, What and How Much
Ned English, *NORC at the University of Chicago*
Ilana Ventura, *NORC at the University of Chicago*
Ipek Bilgen, *NORC at the University of Chicago*
Michael Stern, *NORC at the University of Chicago*

Address-based surveys with web components have been seen as replacements for telephone surveys due to potential advantages in cost, coverage, and non-response. Web surveys are known to have higher response propensities among certain groups, despite coverage limitations for households lacking internet access or proficiency. As a result, it can be challenging to recruit general population households to a web survey from a random sample of addresses. Our paper presents results from multiple studies at the national and state levels where households were contacted via physical mail to answer web instruments. Our analysis focuses on spatial modeling within geographic information systems (GIS) to understand how the kinds of people who respond to web surveys at as recruited by either mode of contact may differ from other modes as well as the population at-large by location and socioeconomic environment. In so doing, we consider the types of households respond to a given mode, their area-level characteristics, in addition to the relationship between geography and coverage or non-response bias. Our paper continues a program of research that uses GIS to approach issues in survey methodology from a different perspective, and is useful to multi-mode ABS and web practitioners.

Inserting Unmanned Aircraft into the Applied Research Process

Joe Eyerman, *RTI International*
Katrina Ladd, *RTI International*
Jonathan Evans, *RTI International*
Mark Bruhn, *RTI International*
Charlie Knott, *RTI International*
Ryan Gordon, *RTI International*
Karol Krotki, *RTI International*
Safaa Amer, *RTI International*

Small Unmanned Aircraft Systems (drones) are quickly being developed for research, public service, and commercial applications across the globe. They are inexpensive, easy to operate, and readily available. Furthermore, they can carry a payload of light, inexpensive, and off-the-shelf sensors that can be used to support a wide range of research efforts. In 2012 RTI launched an initiative to examine the potential to use drones as a tool to support social and behavioral research. The initiative includes projects to examine safety and ethical issues, public perceptions, work and training requirements, policy and regulations, public health applications, environmental monitoring, and survey research. As part of this effort, RTI has collaborated with our university partners to develop procedures and methods for conducting feasibility and field tests of drone based research applications in the United States, Ecuador, and Guatemala. This presentation will discuss the role of the UAS in the research process, the testing process used for selecting the correct combination of aircraft and sensor, the protocol development for safe and ethical applications, and the data management, analysis, and reporting requirements. The presentation will include results from field tests conducted in Ecuador in 2015 and a discussion of pilot studies planned for Guatemala in 2016.

The Value of Self-reported Frequently Visited Addresses in GPS Assisted Travel Surveys

Timothy Michalowski, *Abt SRBI*
Dara Seidl, *Abt SRBI*

Abt SRBI has extensive experience in the application of GPS technology for Household Travel Surveys (HTS) through deployment of 10,000+ GPS devices for various projects throughout the USA. Households participating in GPS travel surveys are deployed small personal GPS loggers for each eligible household member. Participants are instructed to carry the GPS loggers for all travel for 3 to 7 days. As part of the normal industry standard in travel survey recruitment, participants are asked for "frequent travel locations" for all household members. Locations such as home, work, school, and shopping are requested during the recruitment questionnaire. The use of GPS technology allows for high-precision passive travel data collection as an improvement to reliance on traditional self-reported diary methods and corresponding recruitment efforts. With the high precision and frequency of travel location data provided directly by GPS capture, the role of frequent location capture in recruitment efforts is unclear. The frequent locations from travel survey recruitment are geocoded and compared to GPS travel data in Geographic Information Systems (GIS) using spatial statistics. This analysis answers the question, do the frequent locations reported in recruitment enhance GPS travel data? The results of this study inform the level to which the respondent burden of traditional address location reporting is warranted in consideration of rapidly improving GPS technology and analysis capabilities.

Exposing Hidden Structures: Targeted LIDAR Use to Support In-office Review and Validation of an Address Frame
John Liadis, *U.S. Census Bureau*

Assuring an accurate and complete address list for the United States and Puerto Rico is a critical step in the Census Bureau's planning and conducting of each decennial census. Analysis of address and land use/land cover data from a variety of sources indicate in-office methods and data-driven decision-making can substitute for expensive in-field address canvassing in many areas of the country. As a result, and as part of its reengineered address canvassing for the 2020 Census, the Census Bureau is implementing a nation-wide imagery-based review starting in October 2015, to validate the accuracy and completeness of the address list for most of the housing units in the nation. Quality of imagery, however, constrains the review, particularly where tree-cover or cloud-cover obscures the ability to see housing units. In this presentation, we report on the exploratory use of LIDAR to help identify structure locations and street feature not visible through imagery.

Virtual Canvassing: In-office Methods for Validating the Census Bureau's Address List for the 2020 Census
Michael Ratcliffe, *U.S. Census Bureau*

Assuring an accurate and complete address list for the United States and Puerto Rico is a critical step in the Census Bureau's planning and conducting of each decennial census. For the 2010 census, the Census Bureau conducted a full address canvassing operation, with field workers traversing nearly every road in the nation to verify and update addresses in its Master Address File (MAF). Analysis of address and land use/land cover data from a variety of sources indicate that in-office review methodologies and data-driven decision-making can substitute for expensive in-field canvassing operations. As a result, starting in October 2015 as part of its reengineered address canvassing for the 2020 census, the Census Bureau is implementing in-office methodologies to validate the accuracy and completeness of the address list for most of the housing units in the nation. In this presentation, I report on the use of address, land use/land cover, and imagery-based data to evaluate and validate the completeness of information for addresses in the MAF for small geographic areas throughout the United States.

Longitudinal Data Collection Issues

The Development of a Research Portal of Longitudinal Studies in the UK Utilzing Standard Structured Metadata
Jon David Johnson, *Centre for Longitudinal Studies, UCL*

The UK is home to the world's largest and longest-running longitudinal studies. Cohort & Longitudinal Studies Enhancement Resources (CLOSER) aims to maximize their use, value, and impact both at home and abroad. Bringing together eight leading studies, the British Library and the UK Data Service, CLOSER works to stimulate interdisciplinary research, develop shared resources, provide training, and share expertise. At the heart of the project is the assembly of a metadata repository encompassing the full range of information about the studies, from study description to data collection methodology, questionnaire capture and data. Layered on top of this are a controlled vocabulary and other contextual information. The project has very specific challenges. The studies are based in six organisations geographically dispersed across England. Leadership of the studies crosses scientific domains (some primarily medical, some social science focused) and the longevity of the studies (starting

as early as 1946) means that the information available is inconsistently catalogued and held in various formats. Utilising the DDI-Lifecycle metadata standard, the project is creating a research discovery portal that will be interoperable with similar efforts on other similar studies to enable discovery and comparison of questions and data from these studies. The project has primarily used off-the shelf software, combined with some lightweight tools this has enabled mass ingest of metadata into a repository in a consistent and manageable process. The presentation will focus on the challenges of creating such a rich metadata repository, drawing out the main lessons learnt, particularly in the areas of discovery in a multi-disciplinary research environment, and the complexity of harmonisation at scale across these and similar longitudinal multi-purpose studies.

Using Metadata Standards to Document a Multi-disciplinary Longitudinal Study of Aging Called MIDUS
Barry T. Radler, *University of Wisconsin-Madison*

Adhering to research data management principles greatly clarifies the processes used to capture and produce datasets, and the resultant rich metadata provides users of those datasets the information needed to analyze, interpret, and preserve them. These principles are even more important with longitudinal studies that contain thousands of variables and many different data types. MIDUS (Midlife in the United States) is a national longitudinal study of approximately 12,000 Americans that studies aging as an integrated bio-psychosocial process. MIDUS has a broad and unique blend of social, health, and biomarker data collected over 20 years through a variety of modes. For nearly 10 years, MIDUS has relied on a technological metadata standard called DDI to help manage and document these complex research data. DDI provides structured metadata that make datasets machine-actionable, meaning that computer networks and systems can interact dynamically with the dataset elements, thereby facilitating search, discovery, harmonization, analysis, and interpretation. In late 2013, the National Institute on Aging funded MIDUS to improve its extant DDI portal infrastructure by creating a DDI-based, harmonized data extraction system. Such a system allows researchers to easily create documented and citable data extracts that are directly related to their research questions and allows more time to be spent analyzing data instead of managing it. This presentation will explain the rationale and methods of the project, with a demonstration of the resultant MIDUS DDI portal.

Maximizing Test Data on a Longitudinal Study
Catherine E. Billington, *Westat*
Laura Branden, *Westat*

The quality of test data directly affects the quality of the final data. A case study of the National Health and Aging Trends Study (NHATS) argues in favor of maintaining a test data set for longitudinal studies with an annual data collection. NHATS gathers information annually in person from a nationally representative sample of Medicare beneficiaries ages 65 and older. NHATS presents a particularly challenging set of testing and training requirements. Since data are pulled and converted from multiple sources, test data must be available while data collection is in process in order to meet the project schedule for training and fielding for the next round. NHATS project staff sustains a test database that parallels production data and is maintained round-to-round. Test cases from the previous round are closed out according to detailed specifications that anticipate testing and training needs. Detailed scenarios are constructed to assess complex programming and specifications changes. Test data are reviewed and edited to provide sufficient variance for testing and training. Test data provide preload files for testing all instruments including the interview management system, the field operating system, the editing system, and paradata. Experienced testing staff receives updated training every round. This

approach allows researchers to sustain a tight production schedule and creates realistic data that support both complex testing and training requirements and provides a model for other longitudinal studies.

Can We Adjust for the Rubber Band Effect? Using Aggregated Survey Data to Model Lag Effects in CAHPS Surveys
Mark W. Andrews, *Ipsos Public Affairs*

Modeling lag effects in non-panel longitudinal survey data offers a challenge to researcher particularly when the data is highly correlated. The first step of this analysis is to aggregate to specific subpopulations in the data and thus use the subpopulations as the unit of analysis. A simple lag effect can then be defined as the predictive power of a change in variable X between T1 and T2 in determining a change in variable Y from T2 to T3. Unfortunately, this model falls apart when the variables X and Y are highly correlated resulting in what we will refer to as the rubber band effect. As anyone who has worked on tracking studies knows, large fluctuations in results are more often than not followed by reverse movement in the following period as it springs back to the true trend line. A change, therefor, from T1 to T2 is predictive of a reverse change in T2 to T3 for variable X and this negative predictive power is also passed to variable Y because they are correlated and both changes are measured from the same sample. It is the paper to examine this problem using Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey data which are large databases with distinct subpopulation markers and highly correlated data. This paper will discuss the literature on this subject and follow up with sample simulations in order to test available models and/or to determine if any rules of thumb can be developed to adjust for this effect. We will then take the lessons learned and apply it back to the CAHPS survey data in an attempt to accurately model lag effects using real member/patient satisfaction data.

Racial Attitudes in the U.S. and Abroad

Racial Disparities in Views of New Orleans' Recovery Ten Years After Hurricane Katrina
Jamie Firth, *The Henry J. Kaiser Family Foundation*
Elizabeth Hamel, *The Henry J. Kaiser Family Foundation*
Mollyann Brodie, *The Henry J. Kaiser Family Foundation*

Ten years after Hurricane Katrina devastated the Gulf Coast, The Henry J. Kaiser Family Foundation released its fourth survey of New Orleans residents since the storm. The series of surveys (conducted in 2006, 2008, 2010, and 2015) paints a representative picture of the experiences of New Orleans residents throughout 10 years of recovery and rebuilding. Even changes in the survey methodology over time highlight the effects of recovery. In 2006, geographic area probability sampling was used and interviews were conducted face-to-face; by 2015 telephone coverage and available Census data made RDD landline and cell phone sampling possible. The surveys show that New Orleans has made remarkable progress since the storm in many areas, but that progress has been unevenly felt along racial lines. Racial disparities are not a new problem in New Orleans, yet a concerning trend is the widening of the racial gap in some areas. For example, in 2006 a majority of whites and African Americans felt it was a bad time for children to be growing up in New Orleans; by 2015, 70 percent of whites say it's a good time for children to grow up there, compared with 37 percent of African Americans. Reports of neighborhood amenities have improved over time for both groups, but African Americans remain less likely than whites to say their neighborhood has sufficient places to buy

groceries, restaurants, and places for children to play outside. Racial differences exist in perceptions of neighborhood safety and susceptibility to crime as well. These results provide a case study of what may happen in disaster recovery, when those who were disadvantaged from the outset are in danger of feeling left behind in the efforts to rebuild.

Pulling the Curtain Behind Anti-Americanism in Russia: An Examination of Two Theoretical Approaches

Natalie Manaeva Rice, *University of Tennessee*
Dean P. Rice, *University of Tennessee*
Howard L. Hall, *University of Tennessee*
Oleg Manaev, *University of Tennessee*
Michael Fitzgerald, *University of Tennessee*
Yury Drakakhrust, *Radio Free Europe / Radio Liberty*
Maryna Bahdanovich, *Freelance*

From 2011 to 2015, tensions and animosity between Russia and the United States were increasingly reminiscent of the worst years of the Cold War. Possibly the most bothersome aspect of this rise in anti-Americanism was that the hostility of the Kremlin was so strongly reflected by and supported by the public opinion (Fisher, 2015). Public opinion polls showed the negative attitudes of the Russian public towards the United States reached a peak not seen since the collapse of the Soviet Union. This dramatic rise of anti-American sentiments calls for an in-depth investigation of the nature of anti-Americanism in contemporary Russia. Two alternative viewpoints predominate within the literature on anti-Americanism. The first, represented in works of Hollander, Revel, Rubin and Rubin and others, describe anti-Americanism as a syndrome emerging from a cultural predisposition. The second view, which Chiozza (2010) called the Dimensions of America theory, insists that foreign anti-Americanism is not a closed cognitive structure, but rather an "aggregation of considerations, predispositions and information." Our statistical analysis of survey data from Pew Research Center, conducted in Russia in 2012, investigated the nature of Russian anti-Americanism and placed it in the context of two alternative viewpoints on anti-Americanism. Results from cluster analysis were especially helpful in addressing this issue. Data analysis produced a somewhat paradoxical result: anti-Americanism in Russia seemed to fit both approaches simultaneously depending on the grouping of respondents. For a significant part of the population anti-Americanism was a deep seated cultural trait, an ideological lens that affected their perception of everything American. Sentiments of the other part of the population, however, seemed to be more complex: these respondents simultaneously held different opinions towards different aspects of the United States and its influence. This research demonstrated that both approaches to anti-Americanism can be applicable simultaneously within a single country.

Partisanship, Local Context and Attitudes toward Immigration and Refugee Policy: Evidence from Canada

Timothy B. Gravelle, *University of Essex*
Keith Neuman, *The Environics Institute*

The 2015 Canadian federal election campaign brought to the fore partisan cleavages in approaches to immigration and naturalization policy, refugee policy, and multiculturalism. At the level of mass public opinion, recent research on immigration attitudes in Canada and elsewhere has pointed to a variety of factors to explain individual-level attitudes in these domains. These include party identification, economic interests, and feelings of cultural threat. Additionally, in the American context,

there is a growing literature on the effects of local demographic (specifically ethnic) context in shaping attitudes toward immigration, sometimes interacting with party identification, thus highlighting the heterogeneous effects of local context on different segments of the public. Such a contextually-oriented approach, however, has been pursued by relatively few analysts of Canadian public opinion. This paper endeavors to fill this gap. It brings together data from the 2015 Environics Institute Focus Canada survey, the 2001 Census, and enhanced data from the 2011 National Household Survey to answer the questions of what leads Canadians to hold open or restrictionist attitudes toward immigrants and refugees, focusing on the roles of party identification, contextual measures of local immigrant, visible minority, and religious minority populations, and the interactions of party identification with contextual measures.

Adapting the Alternative Questionnaire Experiment for a Telephone Survey: Preparing for Changes to the 2020 Census
Eileen Patten, *Pew Research Center*
Anna Brown, *Pew Research Center*
Kim Parker, *Pew Research Center*

In preparation for the 2020 decennial census, the U.S. Census Bureau has been conducting experiments testing different ways to ask about race and ethnicity, in part to address high shares reporting "some other race" and high refusal rates for the race item, especially among Hispanics. While no final decision has been made, results from these experiments suggest that race and Hispanic origin could be combined into a single question (from its current two-question format) without lowering the population estimates of different race groups or Hispanics. In anticipation of this possible change, Pew Research Center conducted an experiment to test how to effectively implement the potential new race item on a telephone survey. This change will be essential for telephone surveys so they can collect comparable data on race/ethnicity to be used in weighting samples to national parameters established by Census Bureau surveys. Using the American Trends Panel, we separated respondents into three groups: phone respondents with response categories read aloud for all seven race and origin options in the census question, phone respondents with an abridged list of response options read aloud that more closely mirrors our current telephone survey race question, and web respondents with a question and mode comparable to what would be included on a Census Bureau survey. Across major racial and ethnic groups—Hispanics and non-Hispanic whites, blacks and Asians—the shares for the phone survey more closely matched the web version when the long list of response options was read. As the Census Bureau found in their experiments, the new question tested on the phone resulted in a much lower share reporting "some other race" (6% vs. <0.5%) and lower refusals (1.9% vs. <1%) compared with the standard two-question race format.

Problematizing the Construct Validity of Survey Questions: Findings from Cognitive Interview Testing of Questions about How Often Physicians Consider Race/Ethnicity and Other Cultural Factors When Providing Healthcare Services
Marko Salvaggio, *National Center for Health Statistics*
Sheba K. Dunston, *National Center for Health Statistics*

Cognitive interview methodology is a qualitative question evaluation method that examines the construct validity of survey questions. Guided by cognitive interview methodology, this presentation discusses research findings that problematize the construct validity of four survey questions that ask physicians how often they consider race/ethnicity and other cultural factors when providing healthcare

services. Each question was about a particular healthcare service: (1) assessing patients' medical needs, (2) diagnosing patients, (3) treating patients, and (4) conducting health education with patients. Based on qualitative data collected from 20 cognitive interviews with physicians, we explore how physicians interpreted the major constructs of each question and formulated their responses. We found that physicians were consistent in their interpretation of the major constructs of the survey questions, suggesting that they understood each of the healthcare services in a similar way. They were also consistent in their interpretation of race/ethnicity. Physicians varied in their interpretation of other cultural factors. When formulating their responses to each question, some physicians had difficulty in separating each healthcare service and thought of them as part-and-parcel to an overall patient assessment. They also had difficulty providing a universal answer about their patients because they felt that patient assessments are situational. Further, physicians varied in their perceptions about the potential for race/ethnicity and other cultural factors to influence how they provide one or more healthcare service. Accordingly, while physicians tend to interpret the question constructs in a similar way, their perceptions about race/ethnicity and other cultural factors potentially have far greater effects on the construct validity of the questions than mere interpretations of the constructs. Questionnaire design and evaluation can benefit from our findings, which problematize the notion that construct validity is determined by the degree to which a construct measures what it claims to measure based on respondents' interpretation of that construct.

Impact of Interviewers and Sample Frame in International Contexts, sponsored by WAPOR

Mobile and Landline distributions in Western Europe: A Comparison of Phone User Status for 20 Western European Countries

Ken Kluch, *The Gallup Organization*
Sofia Kluch, *The Gallup Organization*

Utilizing data from the Gallup World Poll, this paper explores three elements of dual frames designs in Western Europe general population research; representativeness of the samples to the population census, differences in opinion based on self-identified status, and demographic differences of the both frames and self-identified phone users. The exact mobile and landline sample distribution varies widely across studies and requires constant evaluation and adjustment to ensure that the overall sample provides an accurate and balanced representation of important demographic groups. Data was analyzed in comparison to census demographics, as well as call metrics for landline and mobile samples. Respondents were asked about their preferred phone use for personal telephone calls. Based on their answers, respondents were classified as mobile only, mobile mostly, use both equally, landline mostly or landline only. Respondents' answers to opinion related topics on leadership approval, life evaluation, current economic conditions, and confidence in the government were analyzed to determine potential differences in topical issues based on phone user status. Study populations were also examined by demographic characteristic to determine differences among phone user groups. The sample findings illustrate the need, or lack thereof, for dual frames and identify which demographics groups are most vulnerable to underrepresentation in these sample types. While some countries maintain a healthy balance of landline and mobile phone ownership (according to the Eurobarometer), their populations are more uniform, with less dramatic differences in the demographic characteristics. In other countries there are more pronounced differences that require a more strategic balance of landline and mobile samples. Findings among user status suggest that in addition to the expected demographic differences, there are substantial differences in opinions among these user-types, suggesting the need for each

group's inclusion in specific country samples and the readiness of other countries to move to a mobile only sample.

Toward a Better Understanding of Interviewer Effects in a Nationally Representative Survey in Tunisia

Zeina Mneimneh, *University of Michigan*
Julie de Jong, *University of Michigan*
Mansoor Moaddel, *University of Maryland*

Research has shown that interviewers can have important effects on respondent answers. Potential bias introduced by interviewer gender and religious wardrobe on related survey items is of particular concern in the gender-segregated and religious context of the Middle East and North Africa. For example, studies in the region have found that male respondents reported more egalitarian views to female interviewers (Benstead, 2013) and that interviewers wearing Islamic (rather than secular) symbols and Islamic hijab (vs. no hijab) received increased reporting of religious attitudes either directly or through an interaction with respondents characteristics (Blaydes & Gillum, 2013; Benstead, 2014; Koker, 2009; Mneimneh et al., 2015). Moreover, we have recently shown that an interviewer's own religious attitudes affected respondent's reported religious attitudes independent of interviewer religious wardrobe. The effect of an interviewer's attitudes was as large as, and sometimes larger than, the effect of the interviewer's religious wardrobe (Mneimneh et al., 2015). The literature, however, is lacking on an explanation of the mechanism of these effects. Are interviewers mirroring the attitudes of the respondents they are interviewing or are they projecting their own attitudes on the respondents? Are the effects transmitted through potential side conversations about religious topics between the respondent and the interviewer? Using recently available panel survey data from a second wave of data collected in Tunisia in 2015, this paper investigates these research questions by looking at interviewer's attitudinal measures collected before the field work and contrasting their effects with interviewer measures collected after the field work. Moreover, observational measures on side conversations related to religious and political topics were collected, allowing for investigation of the potential mediating or moderating effects on the relationship between interviewer's and respondent's attitudes.

How do Interviewer Attitudes, Personality and Interpersonal Skills Affect Response Rates? Evidence from a Panel Survey in India

Charles Q. Lau, *RTI International*
Ashish Kumar Gupta, *TNS/India*
Devra Moehler, *Facebook*
Andrew Fiore, *Facebook*

Recent research demonstrates that interviewer attitudes, personality, and interpersonal skills affect response rates in face-to-face surveys (Blom and Korbmacher, 2013; Durrant et al., 2010). For example, less assertive, more extroverted, and more open interviewers obtain higher response rates in Britain (Jäckle et al., 2013). Yet there is little understanding of how interviewer characteristics operate in different cultural contexts. In India, for example, cultural norms governing a survey request are different: there is a greater emphasis on establishing a personal connection with potential respondents and less emphasis on transactional elements such as incentives and reciprocity. As a result, a different set of "soft" interviewer characteristics may affect respondent cooperation. Further, the optimal set of interviewer characteristics may also depend on whether the potential respondent is male or female. We investigate the effects of interviewer experience, attitudes, personality, and interpersonal skills on

response rates in the second wave of an in-person panel survey (n = 5,167) based in the Indian state of Uttarakhand. Our analysis benefits from having an interpenetrated design, in which interviewers are randomly assigned to sample cases. This design allows us to overcome a common hurdle in interviewer effects research (the inability to distinguish interviewer effects from area effects) and allows us to correctly estimate the variance attributable to interviewers. In the analysis, we focus on two outcomes: (1) locating the respondent, and (2) whether the respondent agrees to participate, conditional on being located. Our logistic regression models include respondent demographics (age, gender, education, literacy, tenure in home, caste), interviewer characteristics (demographics, experience, attitudes, personality, and interpersonal skills), and area fixed effects. Guided by previous research, we will include interactions between respondent and interviewer characteristics. This analysis will highlight how interviewer characteristics affect non-response errors in different cultures and will provide insight into interviewer recruitment, training, and supervision.

Split between Mobile and Landline Interviews in Dual Frame Telephone Surveys in International Research
Hayk Gyuzalyan, *TNS Opinion*

One of the big issues in the telephone survey design is the allocation of the mobile and landline samples, and possibly whether to use dual frame at all. This aspect of the survey design becomes more complicated in the context of international telephone research. Dual frame research offers a better population coverage for telephone surveys at the expense of introducing the differential response between modes, mode effects for completing over mobile and landline, and different selection modes between landline (usually household) and mobile (usually single owner). The practical solutions to depend on several factors. Firstly, the split between four populations in a country: landline owners, mobile owners, both mobile and landline owners and people not having either. Another factor is the relative cost of mobile and landline interviews. And finally, availability of public directories, development of Robinson lists, and in some countries, regulation of Random Digit Dialling, are a big factor affecting the design of dual frame telephone samples. All of these factors have been changing rapidly over the last few years, practically every year. The survey practitioner needs to decide how the most recent changes will offer the best solution to the survey aims and objectives and find the optimal survey design. In this paper we analyse the response outcomes of 24 Flash Eurobarometer surveys among the general public in 28 countries of EU, conducted between 2011 and 2015. We use the data to analyse how the distribution of contact outcomes have changed in that time period for landline and mobile interviews. The analysis is based on the comparative analysis of contact outcomes by socio-demographic groups and achieved demographic profile between the two modes of data collection. In conclusion, we look at the likely development of the trend in the near future.

Measuring and Classifying Environmental Attitudes

Solving the Drought in California
Iris Hui, *Stanford University*
Bruce Cain, *Stanford University*

California is experiencing one of the worst droughts in history. The drought problem is due in part to natural causes, including low precipitation and global climate change, and in part to structural problems, such as a lack of adequate water storage, perpetual overuse of groundwater, over-population, over-consumption of water in agriculture. The state is in dire need to create a long-term

policy solution to tackle the water crisis. Water experts are well aware that the water crisis requires a holistic and multipronged policy solution. Yet policies that have been adopted through ballot propositions or legislative bills tend to be ad hoc or compartmentalized in nature. Voters or survey respondents are often asked to evaluate a single policy at a time. Using latent-class analysis on public opinion surveys, we identify politically viable policy bundles and map potential electoral coalitions. We design survey-embedded experiments to test how the latent groups respond to different information and cues over several controversial issues, including consumption of recycled water, water use restriction, infrastructure spending and restriction on agriculture. These experiments aim to identify messages that can provide the largest persuasion effect. Preliminary evidence suggests communication of scientific facts and emphasis on scarcity leave stronger impact, although the persuasion effect varies substantially across issues. In addition, results from latent-class analysis reveals that the correlation between environmentalism and partisanship is not straight-forward. We identify heterogeneous subgroups among self-identified Democrats, Republicans and independents. The results provide a realistic political guideline for policy makers in the state of California.

It's Not My Consensus: Motivated Reasoning and the Sources of Scientific Illiteracy
Josh Pasek, *University of Michigan*

Individuals who provide incorrect answers to scientific knowledge questions have long been considered scientifically illiterate. Yet increasing evidence suggests that motivated reasoning, rather than ignorance, may be responsible for many of these incorrect answers. This paper uses a novel set of survey measures to asses two processes by which motivated reasoning might lead to incorrect personal beliefs: motivated individuals may fail to recognize the presence of a scientific consensus on some issue or they may perceive the existence of consensus while questioning its veracity. Using data from a four-wave panel study of 3,193 Americans drawn from an Internet panel, this study simultaneously examines perceptions of what most scientists would say and personal beliefs about 11 issues for which a scientific consensus exists. Respondents varied in both the scope of their awareness of the scientific consensus as well as their personal beliefs. On average, for any given question, respondents were about 7 percentage points more likely to know the scientific consensus than to believe it. Further, the results revealed that religiosity and partisanship were closely related to both individual-level awareness of the scientific consensus and the extent to which individuals asserted beliefs that contradicted their perceptions of that consensus. Collectively, these pathways almost perfectly predicted the scope of disagreement with the scientific consensus for the issues examined. Notably, however, the relative prevalence of each process (incorrect perceptions of consensus and rejection of a known consensus) appeared to depend on both the scientific issue and the motivational pathway under examination.

Doubt Dries Up: An Analysis of the Effect of Weather on Declining Global Warming Skepticism in the United States
Christopher P. Borick, *Muhlenberg College*
Erick Lachapelle, *University of Montreal*
Barry Rabe, *University of Michigan*

Over the course of the last decade there has been considerable shifts in terms of American acceptance of evidence of global warming. Between 2008 and 2010 numerous surveys indicated a significant increase in the percentage of Americans that did not believe there was solid evidence that global warming was occurring. However, after a few years of fairly stable views on the existence of

global warming there has been a substantial decline in skepticism among Americans between 2014 and 2015. While numerous factors may be responsible for the declining levels of doubt regarding global warming, a widely suspected cause of the shift is experienced weather. According to NASA, 2014 and 2015 are likely to be the hottest years on record and areas of the western United States have experienced an extraordinarily intense drought. Using data from the National Surveys on Energy and the Environment (NSEE) we attempt to isolate the effects of weather on the declining levels of climate change skepticism in the United States. In particular we use geo-coding to link experienced weather with both expressed beliefs regarding the existence of global warming and individual descriptions of the factors that lead to their conclusions about the reality of climate change.

Public Opinion and the Environment: The Nine Types of Americans

Jennifer Benz, *NORC at the University of Chicago*
Emily Alvarez, *NORC at the University of Chicago*
Liz Cantor, *Rutgers University*
Trevor Tompson, *NORC at the University of Chicago*
Anthony Leiserowitz, *Yale School of Forestry & Environmental Studies*
Geoffrey Feinberg, *Yale School of Forestry & Environmental Studies*
Seth Rosenthal, *Yale School of Forestry & Environmental Studies*
Emily Swanson, *The Associated Press*

Common rhetoric posits that Americans are highly polarized in their attitudes toward the environment. A new study, conducted by the Yale School of Forestry & Environmental Studies and The Associated Press-NORC Center for Public Affairs Research, challenges this assumption. We find attitudes about the current environmental issues of the day are not only ideologically motivated, but are rooted in how individuals perceive, interact with, and experience nature. Using a nationally representative online survey of 1,576 American adults, nine segments of the public were identified based on a range of attitudinal measures, including the importance individuals place on environmental protection and what the government's role should be in regulating it, whether there exists an environmental crisis, how individuals see themselves in relation to nature, and how individuals respond when scientific and religious explanations conflict. Americans fall into nine segments each with a unique understanding of the environment, perspectives on key environmental issues of the day, and environmental behaviors. The segments do include the more extreme endpoints on the "green" or environmentally friendly side, such as the "Liberal Greens" who are the most likely of any segment to consider themselves to be environmentalists and who tend to engage in more green behaviors, and the more "brown" or anti-environment, such as the "Conservative Browns" who are the most likely of any segment to say the environmental crisis has been exaggerated. But the data show that the vast majority of Americans fall somewhere in the middle, holding complex and nuanced perspectives on the environment. For example, "Outdoor Browns" spend a lot of time outside in the natural environment, yet few express interest in environmental issues. This paper explores the demographics and sociopolitical characteristics, general environmental attitudes, environmental policy preferences, and environmental behaviors of each of the nine segments that make up the American population.

Do Personality Traits Moderate Recognition of the Scientific Consensus?

Nick Allum, *University of Essex*
Josh Pasek, *University of Michigan*

In recent years, it has become increasingly apparent that some individuals fail to recognize the presence of a scientific consensus when doing so might challenge their religious or partisan identity. This sort of conditional disbelief has been regarded as evidence of both motivated reasoning and social cognition. Yet many people continue to hold accurate beliefs about scientific matters even when partisan and religious factors might lead to cognitive dissonance. In this study we test the hypothesis that personality factors might account for when individuals will recognize the scientific consensus despite dissonance or will fail to do so.

To test whether personality traits might undermine motivated reasoning, we present data from a four-wave panel study of an online sample of more than 3,000 Americans. Respondents were asked a series of eleven questions about whether they thought most scientists would or would not endorse particular claims and their perceptions were coded for accuracy. In line with earlier studies, partisans and religious individuals were disproportionately likely to reject claims about climate change and evolution respectively. We then tested whether psychological attributes including need for cognition and openness to experience might ameliorate partisan and religious differences in perceptions. In contrast to expectations, we found that partisans highest on need for cognition were in fact the most discrepant in their perceptions of climate change and openness to experience did little to mitigate these differences. Differences between the most and least religious individuals also did not shrink at high levels of need for cognition or openness. This suggests that the personality traits identified were not particularly relevant to whether people recognized the presence of a scientific consensus.

May 12, 2016
Poster Session #1

Defining Residential Relocation in a Longitudinal Study of Young Children and Their Parents: Is Distance Moved or Frequency of Moves Related to Mental Health Screening Results?

Sarah Godfrey, *NORC at the University of Chicago*
Jennifer Titus, *NORC at the University of Chicago*

This study explores whether distance moved or frequency of moves is associated with screening positively on mental health assessments in young children. Data were obtained for participants in the longitudinal National Children's Study. Brief Infant Toddler Social Emotional Assessment (BITSEA™) scores for children ages 12 and 30 months and parental Brief Symptom Inventory (BSI®) scores were compared among participants who moved and those who did not move through December 2014. Multivariate logistic regression models using different metrics of residential relocation, distance moved and frequency of moves, were tested to determine whether relocation was associated with mental health screening outcomes. The models controlled for the child's sex and the mother's age, race, ethnicity, income, education level, marital status and language. Children who relocated more often, averaging less than 18 months between moves, were at increased odds of screening positive for mental health issues on the BITSEA at age 30 months compared to children who relocated less frequently (OR: 1.87; 95% CI: 1.11, 3.15). In addition, children of non-white mothers compared to children of white

mothers (OR: 1.90; 95% CI: 1.23, 2.95) and children whose mothers speak a non-English language compared to children of English speakers (OR: 2.02; 95% CI: 1.07, 3.79) were at greater risk of screening positive on the BITSEA at age 30 months. Distance moved was not associated with mental health screening outcomes. These findings suggest the importance of obtaining a comprehensive, detailed address history, in longitudinal studies to allow researchers to control for frequent relocation in analyses of mental health outcomes.

Exploring an Adaptive Design in the National Crime Victimization Survey

Stephanie Zimmer, *RTI International*
Marcus E. Berzofsky, *RTI International*
Andrew Moore, *RTI International*
Michael Planty, *U.S. Bureau of Justice Statistics*
Lynn Langton, *U.S. Bureau of Justice Statistics*

Adaptive designs are used to allocate resources to minimize nonresponse bias. If response rates are high, adaptive designs may not have much impact on bias, but may reduce precision. Through simulation, the effect on bias and precision of estimates for adaptive designs for varying response rates and response mechanisms is explored. Not only will varying response rates be explored but differentials in response rates between subgroups will be considered. We will use R-indicators to determine when the level of differential response within a subgroup poses an increased risk for bias. Furthermore, the mechanism by which nonresponse occurs may impact the adaptive design's ability to mitigate bias. For example, if nonresponse is correlated to the outcome of interest (i.e., missing not at random – MNAR), R-indicators based on known demographics may not reduce bias as intended. The National Crime Victimization Survey (NCVS) design is used as an example of a complex survey design as a basis for the simulation. The NCVS is a household rotating panel survey sponsored by the Bureau of Justice Statistics and conducted by the U.S. Census Bureau to estimate rates of violent and property crime and identify characteristics of the crime victimizations. Historically, the NCVS has had high response rates, but the response rates have been declining in recent years. The Bureau of Justice Statistics wishes to be prepared to have design choices available if the trend continues. Our objective is to quantify the effect of different adaptive designs for different scenarios of response and assess the cost benefit of decreasing overall response rates and precision in order to minimize bias. Preliminary results have found that nonresponse is correlated to victimization status. Therefore, our adaptive design simulations will incorporate a MNAR assumption.

Using Contact Attempt History Data to Determine the Optimal Number of Contacts in Current Population Survey Computer Assisted Personal Interviewing Data Collection

Rachel Bray, *U.S. Census Bureau*
Cynthia Rothhaas, *U.S. Census Bureau*
Courtney Reiser, *U.S. Census Bureau*
Jennifer Tancreto, *U.S. Census Bureau*

As survey costs continue to rise, the Current Population Survey (CPS) is making a concerted effort to manage increasing field costs by seeking efficiencies in data collection efforts. Implementing efficiency measures to control costs may result in detrimental effects on data quality. This research uses data from September 2014 through August 2015 to demonstrate the impact on response rates and sample representativeness resulting from implementing cost saving measures in the CPS contact

strategy, by limiting the number of contact attempts or reducing the data collection period by one day. First, an optimal number of contact attempts is determined. The optimal number of contact attempts is that which achieves efficiency gains while resulting in little to no change in the response rate, sample representativeness, or key estimates. Then, the impacts on sample representativeness resulting from limiting contact attempts to the optimal number are described. The impacts on response rates and sample representativeness resulting from eliminating the last day of data collection are also explored.

Accounting for Social-desirability Bias in Survey Sampling: A Model for Predicting and Calibrating the Direction and Magnitude of Social-desirability Bias

Steven H. Gittelman, *MKTG, INC.*
Victor Lange, *Consultant*
William A. Cook, *Eco-Strategic Advantage*
Susan M. Frede, *Lightspeed GMI*
Paul J. Lavrakas, *Independent Consultant*
Christine Pierce, *Nielsen*
Randall K. Thomas, *GfK Custom Research*

The propensity of respondents to underreport undesirable activities and over-report desirable ones biases our data (Social Desirability Bias - SDB). In order for us to use large scale federal benchmark studies as a standard, we must quantify the impact of SDB. However, efforts to quantify direction and magnitude of SDB to date have been substantially unsuccessful. In 2013, the ARF's Foundation of Quality 2 program implemented a large scale study with 17 sample providers providing opt-in nonprobability samples to an online survey, which ran parallel to a dual-frame phone survey. One of the objectives of this study was to quantify the magnitude of SDB and create a method that might be useful when there are differences in results between different modes of interviews (e.g. online versus telephone) due to SDB. To assist in the objectives of this study, data was compared against highly regarded federal studies that were used as benchmarks. The results indicate that some items in the test had a high degree of divergence between the non-probability samples and benchmarks and therefore, the writers of this paper believed that SDB was present. A separate online study using the benchmarks was conducted to assess the magnitude of SDB and create a model that predicts the magnitude of bias to help researchers establish a factor to correct data. It is our belief that the presence of a viable procedure for correcting estimates of SDB could have far-reaching benefits. When examining the consequences of switching to more cost-effective modes (primarily self-administered/online), it is important that the issue of making proper comparisons does not go overlooked.

Evaluating Survey Administration Mode in Individuals with Physical and Mental Challenges: Results from the Home and Community Based Services Experience of Care Survey Field Test

Christopher Cornelius Pugliese, *American Institutes for Research*
Tandrea Hilliard, *American Institutes for Research*
Coretta Mallery, *American Institutes for Research*
Elizabeth Frentzel, *American Institutes for Research*
Susan Raetzman, *Truven Health Analytics*
Beth Jackson, *Truven Health Analytics*

Interviewer administered surveys have long been viewed as the gold standard for collecting data from individuals with physical and mental challenges. Interviews offer a number of benefits for these

populations, including reduction of cognitive burden, fatigue, and issues related to literacy. However, in-person interviews are expensive and little to no research has compared the effectiveness of different interviewer administered modes – including telephone interviews – for individuals with different types of disabilities. In this study, we compared the effectiveness of phone (CATI) versus in-person (CAPI) interviewer administered survey modes based on response rates, respondent characteristics and quality of care outcomes by disability group. The goal was to determine if phone is an appropriate mode for these populations. Our data come from the field test of the Home and Community-Based Services (HCBS) Experience of Care Survey, which asked Medicaid beneficiaries about their experiences with paid staff that support and/or provide care. The survey was developed for adults that fall into four disability groups: 1) aged and/or physical disability (A/D); 2) intellectual or developmental disability (ID/DD); 3) traumatic brain injury (TBI); or 4) serious mental illness (SMI). The field test included 31 HCBS programs in ten states during 2014-2015 (n= 3,226). We randomly assigned 80% of the sample into phone and 20% into in-person survey administration modes. We found mixed results. While overall the response rate was higher for in-person interviews, response rates were actually higher for phone for the TBI, SMI, and ID/DD groups. Within each disability group, differences in consumer demographics and patterns of response were minimal between modes. CATI respondents reported significantly better overall physical health than CAPI respondents. This suggests that use of CATI may be a viable and cost-effective option for fielding surveys to certain disabled populations, particularly for those with mental challenges.

A New Screening Technique for RDD Surveys? Designing and Executing a Methodologically Sound Response Rate Adjustment Factor Study

Victoria J. Hoverman, *Nielsen*
Robin Gentry, *Nielsen*
Yelena Pens, *Nielsen*
Austin Countryman, *Nielsen*

Due to the development of technologies such as answering machines and caller ID, the continual rise of cell phone only households and declining cooperation rates due to participation fatigue with calls, random digit dial (RDD) telephone surveys have become increasingly more expensive and challenging to conduct. RDD telephone surveys allow for the establishment of rapport and trust between a potential participant and a research group, but this benefit cannot be capitalized upon when the phone is not answered. We developed a new screening technique to identify a sub-group of likely unusable telephone numbers in a sample. The two sources used to better identify unusable or non-working telephone numbers were Cell-WINS, "a real time non-intrusive screening process that accurately identifies inactive telephone numbers within a cellular RDD sample," and the Activity Indicator, a measure that determines how active a phone number has been over the past 10 months. These two indicators taken together allowed for the identification of unusable sample telephone numbers, which accounted for approximately 15% of the total starting sample and produced only 0.1% of total completed surveys. These findings proved to provide a significant improvement in calling efficiency, but required an update of adjustment factors used to assign the probability that a particular telephone number is usable. This presentation will address the procedures necessary to properly conduct a response rate adjustment factor study of unknown usability telephone numbers following the regular calling window of a survey. This is important to survey researchers in order to ensure that their response rates are not systematically overstated in future studies. Furthermore, response rate adjustment factors need to be determined using a scientifically sound methodology, so a framework for achieving this goal will be shared with attendees along with best practices for execution.

Sponsorship Matters: Identity and Privacy Impacts on Survey Cooperation Rates

Kate Terfler Williams, *Nielsen*
Robin Gentry, *Nielsen*
Yelena Pens, *Nielsen*
Darin Harm, *Nielsen*

Research has continued to show that the survey sponsor can impact whether people will cooperate with the research request. In addition, trust, privacy and legal rules have become even larger challenges for those conducting social research. It is important that researchers understand the impacts of sponsorship and legal privacy language on survey response. This research investigates the impact of changes in sponsorship and in privacy language on people's willingness to cooperate with a short survey request across different demographic groups. The data for this test are based on sampled households who were mailed a short paper survey in eight large- and medium-sized U.S. cities. Sample members were randomly assigned to three treatment groups: 1) control group with current sponsorship and without privacy language 2) new sponsorship without privacy language and 3) new sponsorship with privacy language. In this paper, we will discuss the impact of sponsorship on cooperation rate and which demographic groups showed the greatest differences in cooperation rate. Further, findings imply that mitigation for the new privacy language was sufficient, but differentially so for various subgroups.

Adapting Conjoint Techniques to the CATI Environment

Kelsey White, *Survey Sampling International*
Edward Paul Johnson, *Survey Sampling International*
Peter Edwin Booth, *Infosurv*

Choice based conjoint, with its ability to closely mimic realistic decision making scenarios and produce robust part-worth utility scores, is one of the most widely used methods of quantifying preference in modern research. However, conjoint has limited reach due to its reliance on visual real estate that allows respondents to compare products. Sometimes the population of interest may be most accessible over the phone, presenting a complication to using traditional choice based conjoint. We explore an option of a Simple Conjoint (proposed by Jordan Louviere as Best Worst Case 2) that can get results similar to a conjoint but keeps the tasks simple enough for respondents to complete over the phone.

We tested this Simple Conjoint in a real life scenario with a ball bearings supplier*. We obtained more than 200 completed telephone surveys in each of 5 regional markets from a hard to reach business audience. The supplier options presented to respondents had four attributes with three levels each. In the exercise, the interviewer described one option at a time (generated from a balanced design), asking the respondent to specify which attribute of the option described is most appealing and least appealing. Supplier options were never compared against each other directly because it is too hard for a respondent to keep more than one option with its corresponding features in mind at a time.

We show how applying the principles of the Simple Conjoint allowed us to obtain the part-worth utilities needed for analysis, while still achieving an overall cooperation rate over the telephone of 62% and a 96% cooperation rate on the conjoint portion of the survey. We also address the limitations of the Simple Conjoint, including its inability to measure interaction among product sets.

*The end client has been masked, but all data and results are real

Can Social Media Utterances be Transformed into Public Opinion Measures? An Emerging Methodology

Michael Elasmar, *Boston University*
Lei Guo, *Boston University*

The potential of mining social media conversations in order to predict public opinion trends has created mixed reactions among public opinion researchers. While some have attempted to find ways of incorporating the patterns present in social media utterances in their own public opinion predictions, most have totally dismissed this new source of information as being devoid of any scientific merit. Aside from obvious sampling issues, the lack of scientific evidence that utterances on social media definitely reflect beliefs, attitudes and behaviors has been a main impetus for totally dismissing these utterances in traditional public opinion research circles. In fact, those who have so far published studies using social media utterance patterns to predict public opinion have almost exclusively come from outside the public opinion research world, and are neither trained in nor familiar with the public opinion scientific methods of measurement and validation that have been refined for over a century. As a result, their analyses have focused on aggregated patterns that can only be validated if they happen to coincide with trends found in traditional public opinion studies and/or if they predict subsequent public behaviors. Such analyses by social media researchers have so far had widely mixed outcomes. This paper re-conceptualizes the analysis of social media utterances by focusing on the social media utterer as the unit of analysis rather than exclusively analyzing the pattern of utterances. This paper asks: Can social media utterances be transformed into valid and reliable indicators of an utterer's beliefs, attitudes and behaviors? Using a psychometric framework for conceptualizing utterances, the authors report on the development of a methodology for transforming tweets into measures and the challenges they faced when applying the methodology and executing the transformations.

We're Calling through Thick and Thin: Exploring Improvements in Productivity of Landline Samples Using Density Stratification

Carol Pierannunzi, *Centers for Disease Control and Prevention*
Machell Town, *Centers for Disease Control and Prevention*
Trent D. Buskirk, *Marketing Systems Group*
Ashley Hyon, *Marketing Systems Group*
William S. Garvin, *Centers for Disease Control and Prevention*
David S. Malarek, *Marketing Systems Group*

Recently, landline samples have become less efficient, requiring larger samples to achieve a consistent number of completed interviews from previous years. The Behavioral Risk Factor Surveillance System (BRFSS) uses both landline and cell phone samples to produce its annual dataset. The landline portion of the data is decreasing as a percentage of the total number of interviews, but is still essential to produce a representative sample. The BRFSS uses disproportionate stratified sampling (DSS) in which telephone numbers are sampled from two strata (lists) based on the presumed density of known telephone household numbers. In this design, telephone numbers are classified into strata that are either high density (known landline telephone numbers) or medium density (all other possible numbers residing in a 100-series bank that has at least one known landline telephone number) to yield a final sample of residential telephone numbers. Telephone numbers in the high density stratum are sampled at a higher rate than those in the medium density stratum. Specifically, this research will focus on the number of completes produced from each of the density strata, and whether there are significant differences in key outcomes between the two density strata. Six states that have both diverse

populations and large landline samples have been selected (IN, WA, GA, TX, PA, FL) for this research. Using 2014 BRFSS landline samples, comparisons are made by state, density strata and region. The research investigates whether the sampling design could be updated by removing less productive areas from inclusion in the overall sample without loss of statistical accuracy while making the samples more efficient and cost effective. Simplification in dual frame weighting might be indicated from such changes.

Comparing Recruitment for Focus Groups and Friendship Groups: Which Methodology Makes Recruitment Easier?

Kerry Yarkin Levin, *Westat*
Jennifer Anderson, *Westat*
Jocelyn Newsome, *Westat*

Focus groups are an excellent technique to capture users' perceptions, feelings, and suggestions about a topic, product, or issue (Ritchie et al, 2014). The groups are conducted in a controlled, neutral data collection environment. Participants typically do not know each other and researchers purposively recruit a demographically diverse group of individuals (Smith 1972). Recently, a new methodology, called "friendship groups" or "friendship cells," has emerged in the market research area (The Friendship Group Toolkit, 2014). This approach involves recruiting a single "source participant" who in turn recruits friends or acquaintances possessing the characteristic(s) desired for the study. The source participant hosts the group in his or her home. The friendship group methodology has the potential to significantly reduce recruiting costs, since researchers must recruit only one participant instead of many. However, a potential risk of this methodology is that the resulting group(s) may not be demographically diverse. This paper will compare recruiting results from two traditional focus groups and four friendship groups with women ages 20-50. In order to address the potential issue of diversity, our design involved testing two approaches to friendship group recruiting. Half of the friendship groups attempted to mimic the traditional purposive method by asking the source participant to recruit a "racially diverse" group of friends, whereas, in the others, the source participant was simply asked to recruit a group of friends, with no mention of diversity. Results will describe the efficacy and costs of each method. We will discuss the advantages and disadvantages of each approach in terms of costs per complete, ability to successfully recruit target populations, relative diversity of the groups, and show rates. We will also discuss upcoming studies that will address important questions such as the impact of the recruiting methodologies on data quality.

Within-household Selection for Telephone Surveys: Findings from Two Statewide Surveys in Iowa

Jill Wittrock, *University of Northern Iowa*
Mary Losch, *University of Northern Iowa*
Rodney Muilenburg, *University of Northern Iowa*
Ki Park, *University of Northern Iowa*
Neal Pollock, *University of Northern Iowa*

The selection of the respondent within the household is a critical step in ensuring a probability sample speaks to the population under study. There are several possible methods available, and while several of these within-household selection methods have been used for years in surveys, few studies have compared selection methods within the same study. This paper presents the results of a selection method experiment in which the three of the most commonly used methods (i.e., Kish/Rizzo, next birthday, and youngest male/female) were tested during data collection of two statewide telephone

surveys interviewing adult Iowans. Both landline and cell phone samples were collected for a total of 3882 respondents. Respondents in the landline samples were randomly assigned to one of three within-household selection methods. Topics in the first survey focused on perceptions, knowledge, and attitudes toward water quality in Iowa, and the second study included topics on awareness, perceptions, and attitudes toward STEM education in Iowa. These studies were in the field between February and August 2015. Age, gender, and birthday of other adult members in the household were asked as part of the demographics battery and were used to determine the accuracy of the assigned selection method. The selection methods were evaluated based on accuracy, demographic representation, response rates, and refusal rates. Our findings are noteworthy in that they do not conform to previous studies using nationally representative samples, for instance, accuracy was lower for next birthday and youngest male/female compared to Kish/Rizzo, and the response rates for Kish/Rizzo were higher than the other two selection methods. These findings suggest that regional differences may play a factor when determining the preferred within-household selection method for a given telephone study.

Using Paradata to Determine the Optimal Number of Screening Contact Attempts for the National Health and Nutrition Examination Surveys

Tala H. Fakhouri, *Centers for Disease Control and Prevention*
George W. Zipf, *National Center for Health Statistics*
Jeffery Hughes, *National Center for Health Statistics*
Minsum Riddles, *Westat*
Tom Krenzke, *Westat*

Screening eligible respondents for participation in household surveys requires considerable effort and time on the part of field operations and is costly, especially when many attempts to contact a household member are made. Beginning in January of 2015, the National Health and Nutrition Examination Survey began collecting electronic paradata, including electronic records of contact, to better inform the operations of the survey, and to increase sample yield and reduce costs. The purpose of this poster presentation is to compare those who are harder-to-contact to those who are easier-to-contact by demographic characteristics and health outcomes. If these groups are different, the inclusion of the harder-to-contact group may be critical for reducing response-bias, which would justify the added cost, effort, and time associated with repeated contact attempts. Paradata from January 2015 to January 2016 will be used for this analysis. The number of screen contact attempts will be categorized into discrete categories (ex. 1-5, 6-10, and 11-20, and 20+ contact attempts) representing a continuum of "easier-to-contact" to "harder-to-contact" respondents. Differences and similarities between these groups will be studied by various demographic characteristics including age, gender, race and Hispanic origin, household income, and rural urban status. Furthermore, differences in the prevalence of various health outcomes including a previous diagnosis of heart disease, cancer, and diabetes will be examined. Findings from this analysis may help inform the optimal number of screening contact attempts. Specifically, if those who are harder-to-contact are similar to those who are easier-to-contact, any additional contact attempts beyond a set limit may sustain response rates but may not reduce response bias.

To Text, or Not to Text: Effects of Recruitment Strategy on Response Rates to a Survey of Nurse Practitioners

Lydia E. Riley, *American Association of Nurse Practitioners*
Michelle L. Cook, *American Association of Nurse Practitioners*
Tyler Litsch, *American Association of Nurse Practitioners*

The American Association of Nurse Practitioners (AANP) has been conducting survey research since 1985, but responses to AANP surveys have been on the decline; the last national survey only yielded a 6.5% response rate. Given the uniqueness of the nurse practitioner (NP) population, AANP conducted a pilot project to assess the effects of recruitment strategy and survey mode on response rates. Of particular interest to AANP was to find out whether sending a survey invitation via SMS message would yield as high of a response as email or mail. Consent to receive SMS messages from AANP was obtained from 672 NPs either at the 2015 National Conference or online through an email solicitation. An additional sample was pulled from AANP's member database to match the first sample based on key characteristics including membership type and specialty. The samples were randomized into one of eight different experimental groups that varied by recruitment strategy and survey mode. Data collection took place between early September and November 2015. Overall, the experimental group that included initial SMS messaging invitations followed by email invitations yielded the highest response (70.4%). SMS messages were effective in bolstering response rates across all recruitment groups. As a stand-alone recruitment strategy, SMS messages yielded a higher response rate than email or mail (39.6%, 9.9%, and 16.1%, respectively). Those who responded by text message were the most likely to be black or Hispanic, while those who responded by mail were most likely to be white. This pilot study demonstrated that SMS messaging is a viable method for disseminating survey invitations to nurse practitioners, and in particular for targeting certain ethnic groups.

Accounting for Social Desirability Biases: An Exploratory Study of a Sensitive Topic in a Government Setting

Scott Peecksen, *Decision Information Resources, Inc.*
David P. Getman, *Decision Information Resources, Inc.*
Ron McCowan, *Decision Information Resources, Inc.*

An audit on data that grantees use to measure and report on their grant performance has to control for social desirability (SD) biases; otherwise, grantees are likely to underreport risks to data quality. Support for this assumption comes from (1) Lee and Renzetti (1993) who argue that because sensitive topics yield data that can result in respondent sanctions, if a third party accessed the data, respondents tend towards socially desirable responses and (2) research indicating that respondents who participated in a financial audit, compared to respondents who did not, were more likely to underreport financial irregularities (Schneider and Wilner, 1990). For a U.S. Department of Education (ED) audit, data-quality risks were assessed within five areas of the performance reporting system that ED staff and grantees used to measure and report grant performance. The five areas were: (1) measurement design, (2) data collection and reporting guidance, (3) data collection and performance reporting processes, (4) ED staff processing of grantee results, and (5) results dissemination. Findings were shared within ED and so there was potential that they could reflect poorly on participants' performance. To reduce potential SD biases, an exploratory design was implemented that utilized (1) validation strategies with multiple document and data source comparisons, (2) interviews of grantees using neutrally-worded questions to reduce respondents' concerns about how answers could be judged, and (3) a two-stage mapping process that allowed respondents' answers to be checked against known data-quality risk factors

without having to ask respondents directly about them. Results identified risks to data-quality in all five reporting system areas. Some risks would not have been identified if respondents' interview answers had not been compared to additional data sources. This design may have reduced SD biases enough to identify important data-quality risks. This poster will present the design details and overall audit outcomes.

Evaluating the Use of a Representative Sample of OB/GYNs to Increase Response Rates, 1997-2015

Lauren M. Stark, *American College of Obstetricians and Gynecologists*
Michael L. Power, *American College of Obstetricians and Gynecologists*
Jay Schulkin, *American College of Obstetricians and Gynecologists*

Surveying obstetricians and gynecologists on their practice patterns is important to provide information for and on a group who provides an increasing amount of care for women. Like other physicians, OB/GYNs face time constraints and an increasingly demanding work environment which create difficulties in obtaining a high response rate and representative sample (Flanigan, et al. 2008). The Research department at the American College of OB/GYNs has been successful at assessing practice patterns of ACOG Fellows, who account for 94% of US OB/GYNs by surveying a representative subgroup of members (CARN) who agree to take surveys without any incentives. Studies cover a wide range of issues important to physician education, informing guidelines and women's health care in general. We reviewed 75 studies in peer reviewed journals from 1997-2015 where both CARN and Fellows were sampled, to asses if CARN was a representative sample of the larger Fellow population and a successful method of increasing response rates. The studies utilized stratified, randomized sampling through mail, email or mixed modes, and data is analyzed both separately and combined. Although response rates for both CARN and Fellows vary across subject type, CARN response rates have been consistently higher (63% versus 37%, P=.001). We suggest that a commitment to engage in research motivates physician's response rates. Only 8% of papers had response differences not associated with demographic differences; 17% had both demographic and non-demographic differences. Secondary analysis suggest that gender differences in the samples may be an important factor in non-demographic differences. We conclude that CARN is representative of the larger Fellow population and a successful method of increasing response rate, however, recruitment strategies to minimize demographic differences are essential. Future research should consider ways to decrease barriers and make research accessible to physicians to increase response rates.

Improving Children's Audience Measurement in a Fragmented Digital World

Diahanna Post, *Nielsen*
Vera Kurmlavage, *Nielsen*
Kelly Bristol, *Nielsen*
Megan Sever, *Nielsen*
Oana Dan, *Nielsen*
Lindsey Rabhan, *Nielsen*

Measuring children in an online environment is fraught with challenges, from technological to privacy concerns. In a world of big data, cookie pools, and multiple data providers, Nielsen's high-quality panels enable us to accurately identify, model, and calibrate audience measurement. But our measurement accuracy depends on our ability to accurately identify our panelists when they are engaging in content consumption. In addition, compliance can be affected by factors ranging from

technical savviness to cognitive abilities (especially when measuring children). We will share findings from our recent significant upgrade to our online meter prompt about compliance improvements but also about its direct effect on our digital measurement. Nielsen has recently introduced a significant upgrade to the on-screen computer prompt to try to better register children as content consumers. The meter deploys a prompt that asks users to identify themselves from a list of all household members. Since householder names appear in a scrollable "drop-down", some panelists have difficulty selecting their name (e.g., children who cannot read, or adults who do not have the patience to scroll). In order to increase compliance with the prompt, Nielsen undertook an extensive redesign effort. The new design was developed based on quantitative and qualitative feedback from former panelists, experimental collaboration with an academic laboratory, and panel testing. The new features (such as image and color preferences) make engaging with the prompt more enjoyable for children. Test results will show the effects on panelist compliance and highlight the importance of personalization and predictable design when developing an interactive measurement instrument. The prompted data forms the basis for ranking and rating products for advertising and content, and we will share changes in those measures as well.

Can Billing Address Matching Solve Cell Frame Noncoverage in Small-area Surveys?

Nicole Lee, *Abt SRBI*
Tara Merry, *Abt SRBI*
Rachel Martonik, *Abt SRBI*
Stephen Immerwahr, *NYC Department of Health and Mental Hygiene*
Michael Sanderson, *NYC Department of Health and Mental Hygiene*

Cell frame noncoverage in small-area RDD surveys is a growing concern due to cell phone only residents who live in a target area but have a telephone number that is not included in the frame. Options for including these "in-area, out-of-frame" individuals have been limited. New cell sample offerings are now available that include address on the sample record. These new sources make it possible to sample cell phone numbers that have billing addresses in the target area but are outside of the area-defined frame. As an experiment, the 2015 New York City (NYC) Community Health Survey (CHS) included a small sample (n = 150) of non-NYC exchange cell phone numbers matched to NYC billing addresses. (This builds on recommendations we presented in 2015 for using residential mobility as a proxy for in-area, out-of-frame households.) The poster presents a variety of sample performance measures, including working number rate, incidence, and cooperation rates. Among other findings, individuals contacted using non-NYC sample matched to NYC billing addresses were only half as likely to be located in NYC compared to those reached using the NYC cell sample frame. Considerations for use of this sample source are discussed in light of performance and the estimated size of the in-area, out-of-frame population. Recommendations for addressing this type of frame noncoverage are also addressed.

Answering the Siren's Call of the Nationally Representative Sample: Guidelines for Avoiding Data Bias and Recommendations for Selecting the Best Panel for Your Survey

Adam B. Troy, *Consumer Reports National Research Center*
Karen Jaffe, *Consumer Reports National Research Center*
Steven Witten, *Consumer Reports National Research Center*

Selecting an appropriate panel for a survey depends on many factors, namely the study authors' tolerance for measurement error, a desire to project statistics onto a larger population, and study budget. Both the recruitment methodology as well as the mode of participant response (e.g., online, phone, paper and pencil, face to face), play a significant role in whether a particular panel is selected. Nonetheless, the literature is rather limited in guiding researchers on panel selection and the incremental value that each panel design offers (Craig et al., 2013). This study compared 4 popular panel types and examined response data from an identical questionnaire. A survey on American attitudes toward medical privacy was developed by the Consumer Reports National Research Center and sent to four organizations advertising a national panel. Panel 1 was an online probability sample using address-based sampling and provided users with a computer and Internet service if necessary. Panel 2 was an online, non-probability opt-in sample that relied on poststratification weights. Panel 3 was a telephone panel that used a random digit dial generation approach on both landline and cell phones. And, Panel 4 was an online, opt-in sample using quota sampling to approximate the U.S. Census population. Panels were evaluated on the availability of demographic information and consistency of demographic and survey responses with other panels in the study. Chi-square statistics for each variable were calculated along with tests of proportion differences. Demographically, respondents from Panel 4 were significantly younger than those in other panels, while Panel 2 respondents were significantly more likely to be White (non-Hispanic) of have a graduate or professional degree. Only for gender and region of residence did panels not differ. Attitudes related to medical privacy also differed significantly across panels. Implications of these findings on nationally representative panel selection will be discussed.

The Changing Contours of American Daily Life Over the Last Half Century

John Robinson, *University of Maryland*

Certain indicators of the nature and quality of life in American society can be revealed in how our citizens use and feel about their time. Time-diary surveys in the US since 1965 have documented important trends in how American life has been changing over the last half century. Perhaps the most notable transformation has been the reversal of gender roles, with women performing more of the paid work in society and men performing more of the domestic work and child care. At the same time, this has meant a significant increase in the free time for both men and women. While increases in free time may be considered a major accomplishment of the American economy, a notably different perspective is afforded when respondents are asked subjective questions about how much they enjoy these daily activities. As might be expected, US respondents do rate both domestic work and (now particularly) paid work as among their least favorite daily activities, they tend to rate many of their free time activities as notably lower in enjoyment -- particularly those involving use of the mass media, not just television (that now consumes more than half our free time), but newer media like the internet and other IT as well. Instead, higher enjoyment is reported for the personal care activities of eating meals, sleep and to a lesser extent grooming and other personal care. Higher enjoyment is reported for certain free time activities, like fitness, religion and socializing (particularly with one's children), but in general, sleep and meal times rate as more enjoyable than most free time. Most recent diary changes in the new

Millennium have been notably smaller and less hectic than those between 1965 and 2000, but they do involve slight increases in (more pleasurable) sleep and (less pleasurable) TV and computer use.

Utility of Event History Calendars
Antonia Warren, *Westat*
Rose McAloon, *Westat*
Shannan Catalano, *U.S. Bureau of Justice Statistics*

Collecting accurate information about retrospective autobiographical events can be challenged by respondents forgetting events or over-reporting due to telescoping. As a result researchers have used a variety of tools to improve reporting, the most notable being Event History Calendars (EHC). EHCs collect reports and timings of events during a reference period. The calendar is usually structured in a grid with event types as rows and delineations of time as columns. The EHC's design facilitates the use of autobiographical memory structures and pathways which allows respondents to implement sequential, parallel, and top-down retrieval of events (Belli 1998). Although, some studies have evaluated EHC's impact on data quality, more methodological studies are need. The purpose of this study is to evaluate the utility of EHC's in a pilot study of over 4,000 women on rape and sexual assault sponsored by the Bureau of Justice Statistics. The study included a screener that asked respondents if a series of 14 acts, ranging in severity, happened during the last 12 months, collecting the month and year of each incident. The EHC used a semi-structured script to document any instances of five event types during the last 12 months, including family events, job changes, vacations, moves, and other events. It was administered prior to the screener in order to improve recall and help respondents more accurately date past 12 month events. In this study we examine if the amount of event types entered on the calendar effects the response time, amount of missing data, and telescoping in the screening questions. We hypothesize that more event types on the calendar will lead to fewer incidents of telescoping, less missing data on the screening questions, and a reduction in response time. Descriptive and regression analyses will be used to assess the utility of the EHC.

Team Effort: Supporting the Collaborative Needs of Respondent Groups in the Economic Census
Alfred D. Tuttle, *U.S. Census Bureau*

The US Census Bureau's Economic Census (EC) is a self-administered omnibus collection requesting various types of data for companies' individual locations, including details about sales, expenses, employment, assets, and other industry-specific inquiries. Much of the requested data may be drawn from or based on information in company records. For large, complex companies, where job functions and responsibilities are distributed among various specialized workers and administrative units, response to the EC is often a complicated, labor-intensive process requiring many steps and multiple personnel, where the primary respondent acts as a de facto project leader coordinating these efforts. The primary respondent must grasp the scope of the request, identify appropriate company personnel whose help is required, and negotiate with them for their cooperation, along with managing the logistics of communicating, transmitting, compiling, reviewing and submitting the requested data. Other data providers on their teams are expected to comprehend the request, work with data systems to retrieve data, judge their adequacy relative to the request, and map them to the format provided by the coordinator. In the absence of (or lack of awareness of) tools available in the reporting system, coordinators sometimes employ "work-arounds" to communicate and transmit data requests, e.g. reformulating requests in email communications, creating their own spreadsheets to collect data, etc. While such "home-grown" adaptations facilitate survey response, they impose additional burden on

coordinators and providers, and may also introduce the risk that data providers will not have important information needed for accurate response. During re-engineering of the electronic collection system for the EC, the Census Bureau took into account the distributed nature of companies' response processes and developed tools to support their collaborative nature. This paper will describe the research into business survey response processes and the tools we developed to facilitate accurate reporting and minimize burden and error.

The Effect of a Pre-due Date Reminder Letter on Nonresponse in a Business Survey

Aryn D. Hernandez, *U.S. Census Bureau*
Cha-Chi Fan, *U.S. Census Bureau*
Alfred D. Tuttle, *U.S. Census Bureau*

Experiments to evaluate alternative collection methods are rare for establishment surveys. In support of a re-engineering of the 2017 Economic Census, the United States Census Bureau began conducting experiments to evaluate alternative contact strategies. The goals of this effort are to improve the timeliness of response as well as overall response, and to reduce collection costs by reducing the number of nonrespondents re-contacted by more expensive follow-up methods. As the Economic Census is conducted every five years, a series of experiments has been implemented using the Census Bureau's annual and quarterly establishment surveys to test the effectiveness of less-expensive mail strategies. The first of these experiments was incorporated into the Business and Professional Classification Survey (SQ-CLASS), a web-based survey conducted quarterly for the purpose of matriculating "births" (businesses with new or recently reactivated Employer Identification Numbers) into the Census Bureau's business survey frame, and assigning them industry classifications. We conducted an experiment in the third-quarter 2014 SQ-CLASS to test the effectiveness of sending a reminder letter prior to the survey due date. About five thousand nonrespondents were randomly selected to receive a reminder letter about two weeks before the due date, and the others received only the typical post-due-date follow-up. Nonrespondents in the experimental group also received the usual post-due-date follow-up. The experimental group's cumulative response increased relative to the control group following the mailout of the pre-due-date reminder. In this presentation, we will describe the design of the experiment and the methods of analysis, and discuss the results in the context of similar research with demographic surveys and theories about survey response.

Party Polarization and Public Support for Government Spending, 1973 - 2014

Robert W. Oldendick, *University of South Carolina*
Lindsey B. Hendren, *University of South Carolina*

There has been an increasing debate in recent years over the extent to which political parties in the United States have become polarized. One of the central issues in the partisan alignment that has dominated politics in the United States since the New Deal period is government spending. Using data from the General Social Surveys from 1973 to 2014, this paper explores the extent of party differences in support for government spending across a range of topics. The results show that while self-identified Democrats are generally more supportive of government spending than are Republicans, these party differences are not consistent across domains, and the general increase in the differences in net spending support between Democrats and Republicans have also not been consistent across topics. Overall these results are supportive of the view of partisan sorting on the issue of government spending, but there is little to suggest that the American public is becoming polarized on this issue. In sum, it is

clear that the Democratic and Republican parties have established rather distinct and opposing views on the question of government spending and that supporters of the two parties generally recognize and adopt these positions. It is also evident that these partisan differences in support for spending are not uniform across issue areas, with larger differences on social welfare issues that benefit disadvantaged groups; smaller, but important differences on issues that are seen as redistributive, but not entitlements; and areas, such as space exploration and defense for which Republicans are more supportive of government spending than Democrats. While the data on the GSS spending ideas are consistent with the narrative of increased party sorting, there is little in these data to suggest that the American public is becoming polarized on the issue of government spending.

Web Surveys Versus Other Survey Modes: An Updated Meta-analysis Comparing Response Rates

Jessica Wengrzik, *GESIS – Leibniz-Institute for the Social Sciences*
Michael Bosnjak, *GESIS – Leibniz-Institute for the Social Sciences*
Katja Lozar Manfreda, *University of Ljubljana*

The aim of this meta-analysis is to update a previous meta-analysis by Lozar Manfreda et. al. (2008), synthesizing experiments published between 1998 and 2005 on response rate differences between Web and other survey modes. Based on 45 experimental studies identified, Web surveys yielded 11% lower response rates on average compared to all other modes. Furthermore, substantial variability around this average effect size was discovered, which could have been partially explained by the sample recruitment base, the solicitation mode and the number of contacts. Other factors as the type of mode web was compared to, type of target population, sponsorship, publication year or incentives showed no significant influence in this previous study. However, the original study by Lozar Manfreda et. al. (2008) faced a statistical power problem due to small numbers of observations for certain moderator levels. Since 2005, numerous studies carrying information about response rates differences between Web and other survey modes have been published. Consequently, we seek to address the following research questions in our update:How robust are the previous findings over time? Do web surveys still yield lower, the same, or even larger response rates? If we still detect any effects, which factors influence response rate differences? To identify studies published after 2005, we used the same eligibility criteria employed by Lozar Manfreda et. al. (2008) and have identified 94 experimental comparisons altogether, extracted from 57 papers. The overall findings indicate a remarkably robust response rate difference over time (12% lower response rates for Web surveys on average). Moreover, because of the increased number of primary studies, we were able to get more accurate estimates for all moderator variables (type of comparison mode, sample recruitment base, target population, sponsorship, solicitation mode, incentives, number of contacts, publication year) considered. Practical implications and avenues for future research are being discussed.

Respondent Driven Sampling and Injecting Drug Users: Lessons from the National HIV Behavioral Surveillance System

Alicia Novoa, *Texas A&M University*
Lindsay Beth Uberseder, *Texas A&M University*

Survey research has long recognized the utility of respondent-driven sampling (RDS) in surveys of hard to reach populations. In the following paper, we describe how RDS works within the context of the National HIV Behavioral Surveillance System (NHBS), provide insight into best practices for RDS sampling, and draw lessons for future research. The NHBS is conducted in the 20 metropolitan areas

throughout the United States with the highest rates of HIV/AIDS morbidity, and is designed to assist state and local health departments in monitoring risk behaviors amongst three high-risk populations: men who have sex with men (MSM), injecting drug users (IDU) and heterosexuals at risk (HET). In this study, we explore the use of RDS in identifying and recruiting IDU respondents across two different years of data collection and across eight different sites. Specifically, we examine the effects of driving distances, local demographics, and recruitment timing on survey participation and response rates. While our research focuses specifically on the NHBS, our conclusions provide insight into surveys of under-represented and hard-to-reach populations.

Simpson's Paradox in Response Rate Calculations

Stanislav Kolenikov, *Abt SRBI*
Kelly Daley, *Abt SRBI*
Julie Pacer, *Abt SRBI*

We present an example of Simpson's paradox in response rate calculations in a dual frame RDD study. Combining disposition counts from landline and cellphone frames resulted in a lower combined response rate than either the individual landline or cell frame response rates alone. As with any Simpson's paradox, there must be an underlying variable that explains the apparent contradiction. We demonstrate that differential eligibility rates in each frame helps resolve the paradox and provide recommendations for reporting of response rates that are compatible with the AAPOR Definitions and Transparency Initiative when the Simpson paradox arises.

Utilizing a Support Vector Machine Framework to Establish a Confidence Metric for Interpreting Automated Coded Survey Responses from a Natural Language Processing (NLP) Model

Jeremy Biggs, *Mathematica Policy Research*
Andrew L. Hurwitz, *Mathematica Policy Research*
Martha Kovac, *Mathematica Policy Research*
Lisa Schwartz, *Mathematica Policy Research*
Lauren Vollmer, *Mathematica Policy Research*

The use of data science techniques is increasing in survey research. One area where survey research stands to benefit is in utilizing machine learning techniques to automate the coding of survey data. Hurwitz, Biggs, Dovgala, and Kovac (2015) presented a novel approach to coding other-specify survey responses by using a Natural Language Processing model. In this model, supervised learning algorithms were used to automate the coding of other-specify survey responses. The model revealed performance similar to human coders and demonstrated itself to be more cost effective as the amount of data increased. We expand upon this work by presenting an adaptation of an approach introduced by Tibshirani, Walther, and Hasite (2001). The modified methodology applies support vector machines to compute the distance from each response to separating hyperplanes defining each coded category. Utilizing other-specify survey responses from two large telephone surveys, we automate the coding of these responses with a NLP model that utilizes an active learning paradigm. The model then calculates the hyperplane distance for each assigned code and determines which coded responses are farthest from the hyperplane. From this distance computation we derive a confidence metric which states that cases farther away from the hyperplane are less accurate and cases closer to the hyperplane are more accurate. In addition to presenting results for the confidence metric we also present outcomes for the NLP model under the active learning paradigm. Finally, we present a cost-benefit analysis demonstrating

how the confidence metric can result in cost savings for quality assurance coding tasks. This methodological adaptation maximizes efficiency by ensuring coded cases in need of review receive that review and eliminates the bias introduced by a human coder deciding which cases need review. Furthermore, this strategy allows for easy interpretation by survey researchers and leverages data scientific techniques in survey research.

Introduction Breakoffs, Questionnaire Breakoffs and Web Questionnaire Length: A Metastudy

Gregor Čehovin, *University of Ljubljana*
Vasja Vehovar, *University of Ljubljana*

Survey breakoffs denote respondents that start to fill in the questionnaire, but fail to complete it and quit prematurely. This problem requires additional attention because it occurs more often compared to interviewer-administered questionnaires. Studies usually confirm that lengthy questionnaires contribute to lower data quality, but the relation between survey length and breakoff rate often remains unclear, which is our main research focus. To show that different reasons contribute to different types of breakoffs, we separately define breakoffs that occur a) at the survey introduction (introduction breakoffs), and b) breakoffs that happen later at some point during the questionnaire, but still before it is finished (questionnaire breakoffs). This is different from the prevailing reporting only of all types of breakoffs together, which has little practical value, because separate reporting offers a more powerful insight into breakoff prevention. We present practical implications of this approach in a case metastudy of approximately 7,000 web surveys and 2,000,000 responses, where we analyze questionnaire length and its impact on both types of breakoffs. The mean total breakoff rate for our sample is 40% and we show that introduction breakoffs highly dominate (presenting three quarters of all breakoffs). In addition, we found significant correlation between questionnaire length and questionnaire breakoffs, but there is almost no correlation between questionnaire length and breakoffs that occur at the survey introduction. To discuss practical implications, we present a framework of breakoff indicators, where we refer to and extend the AAPOR final disposition code 2.12 "Break-off or partial with insufficient information", in order to separate introduction and questionnaire breakoffs.

A Meta-analysis of Telephone Response Rates for Health Surveys

Mazen A. Sarwar, *University of Nebraska-Lincoln*
Struther L. Van Horn, *Kent State University*
Gregory C. Gibson, *Kent State University*

Response rates for telephone surveys have been declining the last ten years with large survey firms obtaining response rates in the single digit range (Pew 2012). While low response rates may not lead to nonresponse error, research indicates that surveys with low response rates are at greater risk of not representing their target population (Dillman et al. 2014). In addition, due to the difficulty in reaching respondents on the phone, researchers have been forced to invest more time and money into their data collection efforts (Zukin 2015). Despite the importance of response rates, there is little consensus on what is considered an "acceptable" response rate. This is particularly crucial for health surveys, where there is a large amount of interdisciplinary work, published across a wide breadth of journals. This paper explores response rates for peer-reviewed, published, telephone health surveys, attempting to establish a baseline for acceptable telephone response rates. Additionally, this paper explores why certain telephone surveys obtain higher response rates. Utilizing a meta-analytic technique, we examine these variables as potential influences on telephone response rates in health

surveys: topic salience, pre-notification, incentives, answering machine messages, and survey sponsorship. Journal articles from the last ten years (2005-2015) were collected from established databases. This research seeks to provide a clearer picture of where telephone response rates stand for health surveys, and provide researchers with a starting point to incorporate elements into their survey instruments and research practices targeting improved response rates.

Group-Mean-Centering Independent Variables in Multi-level Models Is Dangerous

Jonathan Kelley, *University of Nevada*
M.D.R. Evans, *University of Nevada*
Valerie Lykes, *University of Nevada*
Jennifer Lowman, *University of Nevada*

Multi-level models allow us efficiently to combine data from multiple sites and sources and to correctly estimate both contextual and individual-level effects. But how should we scale the variables we use in them? Group-mean centering of independent variables in multi-level models is widely practiced and widely recommended. For example, in cross-national studies of educational performance, family background is scored as a deviation from the country mean for student's family background. We argue that this is usually a serious mis-specification, introducing bias and random measurement error with all their attendant vices. We examine five diverse examples of "real world" analyses using large, high quality datasets on topics of broad interest in the social sciences. In all of them, consistent with much (but not all) of the technical literature, group-mean centering substantially distorts results. Moreover the distortions are large, substantively important differences pointing towards seriously incorrect interpretations of important social processes. Following the examples, we present a formal analysis showing why group-mean centering produces distorted results. We therefore recommend that group-mean centering be abandoned.

Using Data Science to Explore "Data Science"

Marcos P. Sivitanides, *Texas State University*
Francis A. Mendez Mediavilla, *Texas State University*

In August 2009, The New York Times quoted the chief economist of Google, Hal Varian saying "the sexiest job during the next 10 years will be statisticians." In October 2012, the Harvard Business Review published an article by Thomas Davenport and D. J. Patil titled "Data Scientist: The Sexiest Job of the 21st Century". For a few years now, the statistical community has been discussing the fast emerging fields of data science, data analytics, and big data. Along this discussion have been questions about the role that academic statisticians should play in the development of these fields. Such a role is seen in comparison to that of computer science and operations research, seeing that in academia quite a few well established computer science and operations research programs have recently expanded to include these new fields of study. Social media have become a fast, informal way for people across the globe to communicate and exchange ideas and opinions. It is fair to say that we can find out the latest major events, scientific news, social news and trends in all aspects of life, by analyzing the latest social media postings and feeds. In this paper we set out to find out what people are saying on Tweeter regarding statistics, data science, data analytics and big data analysis. We analyzed the contents of more than 100,000 tweets for terms related to the field of data science, analytics, and statistics. The collection included tweets created between January 1, 2010 and November 10, 2015. A frequency analysis and a network analysis allowed us to reveal associations among terms that identify each of these fields. Some

cliques of particular issues are delineated and shown in contrast to other issues. In all, we found some interesting relationships between key terms.

Reducing Respondent Burden in the ACS: Results from the 2015 CAPI Burden Reduction Field Pilot

Todd R. Hughes, *U.S. Census Bureau*
Robert D. Ashmead, *U.S. Census Bureau*
Eric V. Slud, *U.S. Census Bureau*
Rachael Walsh, *U.S. Census Bureau*

The multi-mode data collection strategy of the American Community Survey (ACS) can be perceived as overly intrusive to some respondents, and the Census Bureau is conducting research to address respondent concerns about the burden associated with the number and type of contact attempts that are made. In August 2015, the Census Bureau conducted a field pilot in the Computer Assisted Personal Interviewing (CAPI) operations of the ACS to evaluate a stopping rule based on a "cumulative burden score." For every contact attempt, a score based on our assessment of the relative burden was assigned. Once the cumulative burden score exceeded the pre-determined threshold, the case was pulled from the active workload so no further attempts could be made. The pilot for the cumulative burden score stopping rule was conducted in roughly one quarter of the areas in which ACS interviewing is conducted. Three experimental treatment groups were used in the pilot: a control group which did not have cases removed (treatment 1), a group for which the burden score was displayed to Field Representatives (FRs) and cases were removed (treatment 2), a group for which the burden score was not displayed to the Field Representatives (FRs) and cases were removed (treatment 3). Treatment groups were randomly assigned by Field Supervisory area. Overall, 4.5% and 4.1% of cases in treatment 2 and 3 respectively, were pulled for exceeding the burden score. Key results related to workload, perceived contact burden, interviewing hours and miles, response rates, and field operations are discussed. We find that overall the cumulative burden score stopping rule was effective at reducing some metrics associated with the burden of the contact efforts in the ACS CAPI operation, while also having some impact on response rates.

Potential Word Ordering Effects of the Recent Childbirth Question on the American Community Survey

Gregory Mills, *U.S. Census Bureau*

Improving data accuracy is an ongoing goal of the American Community Survey (ACS). Through some recent verification work, we became aware of potential discrepancies between data from the recent childbirth question on the ACS and fertility figures from vital statistics, particularly for women in late-childbearing years. We then found some inconsistencies with the word ordering of the ACS question in terms of the time reference, which we hypothesized could be leading respondents to misinterpret the question to be asking about lifetime fertility instead of recent fertility. We also hypothesized that these word order effects would be most apparent in older women, who are most likely to have given birth at some point in the past, but less likely to have recently given birth. Through a multi-modal, dual cross tabulation analysis of recent childbirth responses and presence of a young child in the household crossed by age group, we find that there is likely some over-reporting of recent fertility, particularly among older women.

2015 American Community Survey Mail Contact Strategy Modification Test
Sandra Luckett Clark, *U.S. Census Bureau*

An ongoing goal of many surveys is to increase self-response via Internet or mail in order to reduce expensive nonresponse follow-up operations. The Census Bureau conducted the 2015 American Community Survey (ACS) Mail Contact Strategy Test to evaluate several modifications to mail materials and contact strategies in an effort to improve self-response to the Internet and mail modes. We find that removing the prenotice contact does not improve total self-response. However, removing the prenotice in connection with changing the reminder contact from a post-card to an actionable letter that highlights the login instructions does improve Internet response. We also find that expanding the universe for our third reminder postcard mailing to include nonresponding addresses eligible for our telephone follow-up increases self-response. The Census Bureau implemented some of the options tested in this experiment into ACS production starting in August 2015.

Methods for Detecting Telescoping Error in a Cross-sectional Web Design Survey
Bonnie E. Shook-Sa, *RTI International*
Marcus E. Berzofsky, *RTI International*
Kimberly C. Peterson, *RTI International*
Christine Lindquist, *RTI International*
Christopher Krebs, *RTI International*

To achieve estimates that are representative of a particular time period, survey practitioners rely on respondents to accurately place events in time. However, research shows that placing events in time can be challenging (Groves, et al. 2009). Longitudinal studies mitigate this challenge by utilizing a "bounding" interview, but cross-sectional surveys depend solely on respondents' recall to determine if an event took place within the reference period. Telescoping occurs when respondents erroneously report events that occurred outside of the reference period and often cannot be detected or measured in cross-sectional surveys. In this paper we examine the ability of respondents to place events in time and present a method for detecting erroneously reported events through the use of item replication. The Campus Climate Survey Validation Study (CCSVS), sponsored by the Bureau of Justice Statistics and the Office of Violence Against Women, is a web survey of over 23,000 college students administered in the Spring of 2015 at nine institutions across the United States. Respondents were asked about incidents of unwanted sexual contact that occurred during the 2014-2015 academic year. Respondents who indicated one or more incidents of unwanted sexual contact were then asked to provide details about each incident, including the month in which it occurred. A non-trivial proportion of victims were unable to provide the month in which the incident occurred. Students were also asked to indicate when in their lifetime the most recent incident of unwanted sexual contact occurred. This item was used to identify respondents who reported victimizations that occurred outside the reference period (i.e., respondents who potentially telescoped). We explore the characteristics of respondents and victimizations that influenced the ability of respondents to place victimizations in time and assess the impact of detectable telescoping on resulting survey estimates.

Young Children Living in Poverty: An Examination of Available Measures
Angelina N. KewalRamani, *American Institutes for Research*
Anlan Zhang, *American Institutes for Research*
Sidney Wilkinson-Flicker, *American Institutes for Research*

In 2014, approximately 23 percent of young children under age 6 were living in poverty in the United States (DeNavas-Walt and Proctor, 2015). Prior research has shown that young children living in poverty suffer severe academic and social outcomes, and the severity varies by gender and race/ethnicity. Therefore, it is important to be able to examine differences in poverty status for small subgroups. Different data sources are available for measuring young children living in poverty and each offers advantages and disadvantages. Since 1960, the U.S. Census Bureau has produced poverty estimates based on the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC). This official poverty measure consists of a set of thresholds for families of different sizes and compositions that are compared to before-tax cash income to determine a family's poverty status. Developed in 2009, the Supplemental Poverty Measure (SPM) extends information provided by the official poverty measure by adding to family income the value of benefits from many government programs designed to assist low-income families, subtracting taxes and necessary expenses such as child care costs and medical expenses, and adjusting poverty thresholds for differences in housing costs (Short, 2015). Both the official poverty measure and the SPM are based on CPS data. In addition, the U.S. Census Bureau reports on poverty using American Community Survey (ACS) data, which has a substantially larger sample size and enables more detailed analysis of children living in poverty than the CPS. The CPS and ACS differ in several aspects, including sample size, mode of administration, questionnaire design, and reference period. This paper will examine the different measures available for estimating poverty in young children and present CPS and ACS estimates by child and family characteristics from 2009 to 2014. The analyses will help researchers determine which measure best fits their needs.

The Texas Transportation System: How Good is It?
Chris Simek, *Texas A&M Transportation Institute*

As policy makers find themselves on the brink of passing a multi-year transportation bill that would authorize federal spending of up to $325 billion on road, bridge and rail transit projects for six years, state and local governments are preparing by assessing the needs of various user groups. To help identify how well the Texas transportation system is serving those that use it to move goods and/or services, the Texas A&M Transportation Institute (TTI) recently implemented the Texas Freight Pilot Survey, during which more than 500 Texas establishments were interviewed. Respondents were presented with 12 transportation-related factors and were asked to rank how well the current transportation system is performing across four transportation infrastructures (highways, rail, ports and border crossings). Respondents ranked their system performance satisfaction levels using a numerical scale, which was then converted and expressed in letter-grade rankings for this report. The purpose of this poster will be to summarize the research findings and discuss how these Texas specific findings relate to the provisions of the proposed transportation bill.

Collecting and Reporting Post-graduation Employment Data: One University's (Successful) Efforts to Provide More Accurate and Useful Information than the Feds

Nancy Whelchel, *North Carolina State University*
Suzanne Crockett, *North Carolina State University*

The DOE's federal College Scorecard has been criticized for, among other things, the sources used to identify graduates and income data, and for the inability to drill down to findings for specific academic majors. The NC Common Follow-up System has recently released NC TOWER (Tool for Online Workforce and Education Training) to provide similar information for graduates from NC community colleges and UNC system schools. While NC TOWER is in some ways more robust than the College Scorecard, it too is limited by the employment data it uses. As students and families seek information to help make decisions about not only where to enroll in college but what to major in - - for right or wrong often focusing on short term earnings potential immediately after graduation - - the ability of a college/university to provide accurate, comprehensive information about such outcomes can prove invaluable to both students and the university. NC State University has been administering the Future Plans Survey (FPS) to graduating seniors just prior to commencement since Fall 2010. The FPS collects detailed information from seniors on jobs they have secured, graduate/professional school enrollment, and about professional development opportunities in which they engaged while an undergraduate. In this presentation I will offer detailed comparisons between salary data collected by the College Scorecard, NC TOWER, and the FPS. I will also provide evidence of how the FPS data can be used to attract students and help them engage in professional development activities, by academic departments and support units for program assessment and improvement, by students and employers in negotiating salaries, and in other efforts to enhance student success. This presentation will also identify best practices to help institutions interested in collecting similar information engage their own students and effectively use the data for decision-making.

Evaluating the Effectiveness of Disproportionate Stratification to Oversample Subgroups Given Frame and Population Characteristics

Burton Levine, *RTI International*

Disproportionate stratification is a common method to increase the quantity of respondents in subgroup(s) of interest. In a 2009 article, Graham Kalton identifies three conditions for the effectiveness of disproportionate stratification to produce a sample with substantial efficiency gains as compared to a simple random sample. I develop methodology to quantify the effect of these conditions on the utility of disproportionate stratification. The effective sample size of the population always decrease as the oversample of a subgroup increases. Paradoxically, after initial gains, the effective sample size of the subgroup decreases with an increase of the oversample. I show how to calculate when this occurs. I show the change in effective sample size for the subgroup and the population when disproportionate stratification is applied compared to simple random sampling for different frame and population scenarios. I also use this methodology to inform the optimal way to create the stratification.

How do Self and Proxy Responses Differ in Reporting Patient Experiences with Healthcare?

Nikkilyn Morrison, *Mathematica Policy Research*
Nancy Duda, *Mathematica Policy Research*
Nancy A. Clusen, *Mathematica Policy Research*
Karen Bogen, *Mathematica Policy Research*
Sabrina Rahman, *Mathematica Policy Research*

Researchers often must decide whether to accept data from someone who is responding on behalf of a sample member who cannot respond due to physical or mental limitations or language barriers. Accepting such proxy responses increases response rates, but raises concern that the proxies' responses may not reflect how the sample members would have responded for themselves. For the Consumer Assessment of Healthcare Providers and Systems (CAHPS) surveys, the Agency for Healthcare Research & Quality specifies that proxy responses are not acceptable. This paper uses data from three administrations of a mail-only survey of patients' experience with their primary care provider for the Evaluation of the Comprehensive Primary Care Initiative, sponsored by the Centers for Medicare & Medicaid Services. The questionnaire contains the entire CAHPS Clinician & Group Patient-Centered Medical Home survey and other items. Approximately 70 percent of respondents are Medicare beneficiaries. Most data collections do not have enough cases to assess the implications of proxy respondents, but we received over 150,000 responses, including nearly 7,000 proxy responses. Sample members for whom we receive a completed survey shortly before the reminder mailing may receive another survey that is also completed; about 120 of our proxy responses also have an accompanying self-response. We assess the differences between proxy and self-responses for these cases. We will look at agreement throughout the survey, in addition to looking at specific types of survey items, such as sensitive personal information versus respondents' opinions or non-sensitive factual information. Though we cannot know which response is correct, we will be able to assess how much proxy responses differ from self-responses to make a judgment about proxy exclusion rules and identify the types of items where proxy response might result in a greater difference in response.

Using Prepaid Cell Phone Flag to Increase Respondents within a Targeted Population: Efficiencies and Tradeoffs

Alisha Baines Simon, *Minnesota Department of Health*
Kathleen Thiede Call, *State Health Access Data Assistance Center*
Sarah L. Hagge, *Minnesota Department of Health*
Giovann Alarcon, *State Health Access Data Assistance Center*
David Dutwin, *SSRS*
Karen A. Turner, *State Health Access Data Assistance Center*
Stefan Gildemeister, *Minnesota Department of Health*

As the percentage of cell-phone only users has increased over the past decade, targeting specific populations for telephone surveys has become more complex. While geographic targeting has improved, rate centers often represent entire metropolitan areas, which limits their effectiveness. Sampling techniques for cell phones now include the ability to add a "prepaid" flag to sample after purchase. Early experience with the prepaid cell phone population suggests they are more likely to be non-white, lower-income, have lower levels of educational attainment, rent their home, be cell-phone only households, and not have health insurance. Every two years, the state of Minnesota conducts the Minnesota Health Access Survey (MNHA) a population-based dual frame telephone survey on health

insurance coverage and access. Results of the survey are used to determine the size and provide a description of the uninsured population. The good news is that the uninsured population is shrinking due to the federal Affordable Care Act; however, this population is becoming harder to reach. To increase the likelihood of sampling uninsured Minnesotans in 2015, the MNHA sampled prepaid cell phones at a 2 to 1 ratio to non-prepaid cell phone. This paper explores the tradeoffs between the ability to survey more uninsured Minnesotans and the downside of increased design effect after weighting and increased cost. Firstly, we explore how the prepaid cell sample differed from the non-prepaid sample in demographics, health insurance status, health care utilization, and access to health care. Preliminary results indicate the percentage of uninsured Minnesotans in the prepaid sample is approximately two times that of the non-prepaid sample. Secondly, we calculate the cost and design effect tradeoff of adding a prepaid flag, thus increasing sample with certain characteristics, as compared to not using a prepaid flag, but perhaps needing to increase overall sample size to provide accurate estimates.

Measurement Invariance Between Employed and Unemployed Persons: An Example Using a Multi-item Health Measure
Stefanie Unger, *Institute for Employment Research (IAB)*

Surveys often rely on subjective measures for various concepts such as wellbeing, health or happiness. For these measures it is important that different groups of the population understand the measured concepts and items in the same way. Otherwise conclusions drawn from analyses may be misleading. The SF-12 is a widely used scale to measure health-related quality of life. It consists of 12 questions covering both mental and physical health and health restrictions. It has been used by former research to assess the effect of unemployment on individual health. However, there is some reason to question the validity of the SF-12 for the comparison of employed and unemployed respondents because of some critical questions which refer to work or other regular activities. Employed persons who feel restricted at work due to their health may rate their health lower than unemployed persons of the same health status. While there is research on the intercultural equivalence of the SF-12 (Schulz, 2012), there exists no previous research testing if the SF-12 is equivalent across groups with different employment status. I assess, whether the SF-12 is suited to analyse health effects of (un-)employment using the large German Panel Study Labour Markets and Social Security (PASS) which consists of a benefit recipient sample and a sample from the general population. The household panel has been conducted since 2006 and covers over 8,000 respondents each year. The overrepresentation of benefit recipients who are most often unemployed, allows detailed comparisons between employed and unemployed persons. Assessing measurement invariance between employed and unemployed persons is of utmost importance because it is a prerequisite for drawing correct (causal) conclusions. I use the CFA-framework in order to tackle equivalence of health measures for employed and unemployed persons.

Survey Consistency: PARC it
Gary Langer, *Langer Research Associates*
Margaret Tyson, *Langer Research Associates*
Greg Holyk, *Langer Research Associates*
Chad Kiewiet de Jonge, *Langer Research Associates*
Julie Phelan, *Langer Research Associates*

We present a new knowledge management tool specifically designed for survey research practitioners and sponsors – the Polling Archive, or PARC. This cloud-based software stores, searches

and retrieves survey questionnaires, trended topline results, analyses and associated files, improving survey management, enhancing accessibility and transparency of analytical materials and ensuring consistency of administration in field work. Our presentation builds upon the 2016 AAPOR conference theme, "For probabilistic approaches, it is all about the method," while expanding this concept to all survey modes and methodologies. Whether one employs probability-based, non-probability or qualitative methods, fundamentals hold true, including the need for effective storage and retrieval of research materials and precise replication of questionnaire design for time-trend purposes. The ability to quickly and conveniently assemble all previously asked questions, results or reports on a given topic (and/or for a single sponsor), with single-click access to underlying data, is of clear value in producing comprehensive research analyses. Even more fundamentally, question-level search and retrieval of programmed questionnaires is critical in ensuring consistency in field work. PARC provides for exact carry-forward of question wording, interviewer instructions, response option wording and codings, volunteered responses (coded or verbatim), rotates, scrambles, skips, branches, prods and all other embedded instructions. Finally, in the spirit of AAPOR's Transparency Initiative, PARC enables survey producers to provide access to research materials. At a time of increased sophistication across a range of survey practices, particularly in sampling and computational endeavors, PARC fills a fundamental need in the little-addressed area of survey information management, enhancing reliability, analytical opportunities and disclosure alike.

A Well-defined Community: Using Self-defined Neighborhoods to Examine the Differences Between Perceived and Objective Access to Green Space

Madeleine E.G. Parker, *NYC Department of Housing Preservation and Development*
Po-Ju Tuan, *NYC Department of Housing Preservation and Development*

The neighborhood is frequently used as a framework for understanding opportunities and effects on the urban population. Researchers and policymakers are interested in the impacts of neighborhoods on respondents' well-being. Often overlooked, however, are respondents' own definitions of their neighborhoods or communities; boundaries are frequently simplified and assumed to be equal to census-defined boundaries or political jurisdictions. This simplification can result in inaccurately understanding and representing the situations of urban populations. In particular, when it comes to access to green space, urban residents might be in administrative areas with plenty of green space, or might be objectively near parks, but still may not perceive these green spaces as being close to their homes or within their neighborhoods. The Housing and Neighborhood Study (HANS) addressed this question of perceived access to resources by gathering respondents' own definitions of their neighborhoods and perceived proximity to green space. As part of an in-person interview, respondents were asked to define the community in which they lived by drawing (or describing in words) the boundaries of their neighborhood. Results from the spatial analysis were then digitized and linked to other interview data, including home address and how much respondents agreed or disagreed with the statement, "There is a park or green space that is close to my home." This paper uses statistical analysis to examine whether perceived proximity to green space is related to: the objective distance from a respondent's home to the nearest park or other green area; the amount of green space in a respondent's self-defined neighborhood; or the amount of green space in a respondent's administrative boundary, to better understand the differences between perceived and actual spatial neighborhood characteristics.

Pre-election Surveys Using a Multi-modal Interviewing Strategy

Jacqueline Redman, *Franklin & Marshall College*
Scottie Thompson, *Franklin & Marshall College*
Berwood Yost, *Franklin & Marshall College*

Despite the fact that recent polling in the US has performed no worse than in the past on average, significant misses in some races have pollsters and poll watchers concerned about the future of political polling. Of primary concern is the potential biases created by increasing cell-phone only (CPO) households, declining response rates, and the relationship between non-coverage, non-response and voting. The increasing costs required to conduct methodologically sound pre-election polling is another concern. All of these concerns are forcing pollsters to adopt new methodologies, one of which is the use of mixed-mode surveys. Mixed-mode surveys are used to combine the strengths of multiple methodologies in order to achieve higher response rates, more representative samples, and better response quality. There is also evidence that using a mixed-mode design increases the efficiency of a survey. This poster explores the feasibility of using a mixed-mode design (web and telephone) to conduct voter surveys by presenting a comparison of interviewing efficiency, response rates, sample representativeness, and survey estimates produced by using telephone and web-based data collection strategies. The data for this poster come from surveys conducted using list-based samples of registered Pennsylvania voters during the months of June (n=640), August (n=691), and October (n=677) 2015.

Design and Data Science: Changing the Face of Survey Data Reporting

Rob F. Chew, *RTI International*
Joey Morris, *RTI International*
Martin Duparc, *RTI International*
Mahin Manley, *RTI International*
Thomas Miano, *RTI International*
Gayle S. Bieler, *RTI International*
Anupa Bir, *RTI International*
Cynthia Bland, *RTI international*
Kathleen Considine, *RTI International*

Data science is transforming all fields of research in important ways, and the field of survey research is no exception. For instance, the traditional methods of reporting results through the production of static tables and charts are quickly giving way to modern, web- and mobile-based interactive reporting tools. These tools are important for all consumers of the data, from the general public to policy makers to social science researchers, and they change the paradigm from one of passive consumption to one of active engagement. When we design products that take human needs into account, when we iterate through rapid prototyping with the client, we end up with a product that is dynamic, timely, and encourages users and stakeholders to explore and understand the data in ways not possible with static reports. In the past year, RTI International's Center for Data Science has had an opportunity to work directly with clients on solving problems related to the availability and approachability of data collected from complex sample surveys. This presentation summarizes our experience in creating dynamic, interactive reporting tools for the Bureau of Primary Health Care's Health Center Patient Survey and the New York City Office of Financial Empowerment's Mobile Services Study. The resulting tools use only open source software and are accessible from the clients' own websites. The end result is to meaningfully increase the impact of survey data collected by federal, state, and local governments.

Who is Using the Computer? The Impact of Increasing Prompt Frequency on Data Quality and Panelist Cooperation

Megan Walsh Sever, *Nielen*
Vera Kurmlavage, *Nielsen*
Shelli Kashriel, *Nielsen*
Oana Dan, *Nielsen*

In the era of increasing digital media fragmentation and audience expansion, panel-based online measurement provides a critical basis for assessing the demographic and temporal dimensions of individual media consumption. However, one challenge to panel-based online measurement is identifying who is consuming online media at a given time, especially in households where digital devices are shared among different household members, belonging to different demographic segments. Adults and children in a household often use the same computer, impacting the accurate capture of users' demographic, which is essential for digital audience data quality. To this end, Nielsen's panel-based online measurement often utilizes computer metering software to track panelists' activity by requiring users to identify themselves when they start and stop using the computer. Accuracy of this type of metering solution relies heavily upon panelists' cooperation in informing the metering software whenever users change. However, panelists may forget to log off when they stop using the computer. One way to "catch" user changes is to prompt users to re-identify themselves after a certain period of inactivity; however, this process may increase panelist burden and lead to higher panel turnover. This paper explores what the optimal interval for a prompted measurement tool might be to increase data accuracy while balancing the risk of panelist burden. In particular, we explore differences in panelist cooperation between prompting users to re-identify after 15 minutes vs. 30 minutes of inactivity. We present the results of a test in a subsample of households participating in Nielsen's national sample of computer measurement in which panelists were changed from the current 30-minute prompt to a 15-minute prompt. Preliminary research suggests that the 15 minute interval may better capture user changes (especially between adults and children) without significantly increasing panelist burden and without affecting panelist cooperation.

The Impact of Scale Direction, Alignment and Length on Responses to Rating Scale Questions in a Web Survey

Florian Keusch, *University of Mannheim*
Mingnan Liu, *SurveyMonkey*
Ting Yan, *Westat*

Although rating scale questions are used extensively in surveys and have received much research attention in the literature, the direction of the rating scales has not been as extensively studied as have other design features. Previous research found mean shifts and changes in proportions endorsing various response options under different scale directions, indicating that respondents are more likely to select response options at the start of the scale, regardless of whether the response options are presented in an ascending order (i.e., the scale runs from the positive pole or highest level to the negative pole or lowest level) or descending order (i.e., running from the negative pole or lowest level to the positive pole or highest level). However, it is not quite clear yet, whether this effect is independent of other scale features, such as alignment of the scale (i.e., whether the scale is presented vertically or horizontally) and the number of scale points. Also more research is needed on the influence of scale direction on other properties of the resulting answers (e.g., factor structures, internal

consistency of resulted factors). In two experiments in a nonprobability online panel, respondents were randomly assigned to answer two sets of multi-item batteries (4 items on traditionalism and 4 items on political efficacy) on a descending scale or an ascending scale. Preliminary analyses confirms that the scale direction significantly affects resultant survey responses by pushing answers to the beginning of the scale. We found that scale alignment (vertical vs. horizontal presentation of the scale) and number of scale points (5-point vs. 7-point) does not interact with scale direction but that respondents who speeded through the questionnaire were particularly prone to the scale direction effect. Further analysis will focus on the influence of scale direction on factors structure and internal consistency of the resulted factors.

Surveying American Indian and Alaska Native Parents: Identifying Characteristics of Survey Mode Preference
Kathleen Feeney, *Mathematica Policy Research*
Frank Masters, *Mathematica Policy Research*

With access to the Internet widespread, the research community is increasingly utilizing multi-mode and web-based surveys as a means to reduce costs and respondent burden (Couper 2008; Dillman, Smyth, and Christian 2009). However, little is known about how preexisting cultural norms, demographics, and access to technology may affect respondent mode preferences. Our team previously conducted an experiment that found that the introduction of a web option did not influence the response rates of low-income parents, and that parents' mode preference was not influenced by whether the advance letter offered a choice of web or phone or the advance letter only offered web. In a study of children and families in American Indian and Alaska Native (AI/AN) communities, our team anticipates to survey 800 parents of preschool-age children, offering web and telephone modes. We will examine the response rates overall and by mode, as well as examine age, gender, location, and ethnicity in regard to response mode preference. We will explore associations with respondents' mode selection, examining survey methodology factors such as consistent presentation of available survey mode options in advance materials, as well as non-response follow-up efforts, including reminder letters, emails, and on-site parent meetings between a nationally representative sample of parents with pre-school age children and our current sample. We will also explore information collected by field staff regarding access to reliable internet services in the community and examine reported reasons of mode preference by phone respondents. Exploring these associations in depth can help us to detect patterns in survey mode preference for AI/AN parents with young children and inform future surveys about the benefits of different mode options.

Heaping at Round Numbers on Financial Questions: The Role of Satisficing
Michael S. Gideon, *U.S. Census Bureau*
Joanne W. Hsu, *Federal Reserve Board*
Brooke Helppie McFall, *University of Michigan*

Responses to quantitative financial questions on surveys frequently display strong patterns of heaping at round numbers; that is, numbers ending with one or more zeroes. Heaping is often assumed to be indicative of satisficing and reduced-quality data. This highlights a potential trade-off: respondents may answer questions using strategies that reduce burden but yield lower quality data. This paper uses three studies to examine variation in rounding across questions and individual characteristics. We examine rounding on financial questions in the Survey of Income and Program Participation (SIPP), the Survey of Consumer Finances (SCF) and the Cognitive Economics Study (CogEcon). Using Krosnick's

(1991) theory of satisficing and Tourangeau's four stages of response as our analytical framework, we test whether rounding is consistent with satisficing, question sensitivity, or other response strategies. We define three types of financial questions: knowable questions about aggregated values, knowable questions about single values, and unknowable questions. These types vary in task difficulty at each stage of response, and particularly in the information retrieval and integration stages. Preliminary evidence using SCF and CogEcon is consistent with the hypothesis that round number responses reflect cognitive shortcuts and satisficing. Rounding was more common among respondents with low cognitive ability and motivation, and for more difficult questions. Consulting records, which lowers task difficulty, reduced rounding. Patterns across question topics appeared remarkably similar across the two surveys despite differences in mode, question order, and wording. By comparing these results with SIPP 2008 and the redesigned SIPP 2014, we assess whether differences rounding across questions is topic-specific phenomena (and hence similar across surveys), or influenced by survey design (wording and context). These analyses shed light on ways to improve data quality and reduce respondent burden for questions on household finances.

Can a Mail Survey Reduce In-person Data Collection Costs?
Hohn Hartge, *Westat*
Reanne Townsend, *Westat*
David Cantor, *Westat*

Screening for sub-groups within the general population (e.g., young people; low income) is expensive for in-person collections because of the significant effort required to contact and screen a household. One possible method to cut costs is to screen households using a mail survey (Wagner, et al, 2013). Households that do not meet the eligibility criteria do not have to be visited by field staff. This paper assesses the cost effectiveness of this methodology for a survey screening for females under 50 years old. The National Survey on Health and Safety (NSHS) is an in-person survey of adult women below age 50 in five metropolitan areas. NSHS, sponsored by the Bureau of Justice Statistics, is investigating survey methods to collect data on rape and sexual assault. Prior to field data collection, a brief mail survey was sent to each sampled unit to collect household roster information. The mail package included a $2 incentive to return the roster. After three contacts, the roster response rate was 25% and approximately 10% of the original sample was coded as Postal Non-Deliverable (PND). Approximately 60% of the surveys resulted in coding out the household as ineligible. We will compare the cost of the mail survey to what would have been spent if a field interviewer had visited all sampled units. We will extend the analysis to assess where the break-even point would be with respect to a different level of effort for the mail survey with respect to response rate and variations in the incentive. Finally, we will evaluate the eligibility of addresses where PND was returned for one mailing, but no survey was returned for the other mailing. Interviewers visited these households and provided a final determination on the status of these sampled units.

May 12, 2016
Concurrent Sessions B

Development in Online Panels

Developing and Maintaining an Opt-in Panel for Swedish Opinion Research
Karin Nelsson, *Inizio*

Lars Lyberg, *Stockholm University*
Stefan Petersson, *Inizio*

In Sweden the year 2014 was called the super election year because of the elections to the EU parliament and the Swedish parliament. These events generated a demand for continuous polling. Most of the Swedish survey organizations use initial probability samples resulting in estimates or predictions with overstated confidence levels due to oversampling until certain sample sizes are achieved, nonresponse biases, and other nonsampling errors. One critical aspect in most opinion surveys is timeliness. All large media organizations as well as the political parties demand data within extremely tight timeframes leaving very little room for traditional nonresponse follow-up etc. Another critical aspect is the current demand for data-driven journalism. Obviously such demands must be met by nontraditional survey methods such as panel approaches with short turn-around. In this paper we describe our development, implementation, and maintenance of an opt-in panel, that is eventually supposed to be run entirely by our client, Aftonbladet, the largest newspaper in Sweden. Aftonbladet is part of a large media group, Schibsted, with 7,000 employees in 29 countries. The implementation has so far been successful, and the approach is now being exported to other members of the Schibsted media group. We have developed a process for panel recruitment among web sites associated with Aftonbladet and other Schibsted partners in Sweden. Currently the panel has more than 30,000 members and more than 250 surveys have been conducted. To adjust for representation problems we have used various weighting procedures and also received guidance from a scientific board. Procedures are in place to identify, among other things, satisficing behavior and mischief. Panel data have been calibrated with official statistics. We will present our work procedures and some examples of results obtained from our panel as well as some comparisons with results from other service providers working on election predictions.

Can Using a Mixed Mode Approach Improve the Representativeness and Data Quality in Panel Surveys?

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

A proposed solution to the current challenges associated with individual mode surveys is to recruit panels of respondents who agree to complete a series of surveys over the internet. However, this approach has been criticized given that surveyors do not fully understand the way different segments of society use the web and the homogeneity of samples. Recently, surveyors have explored the use of mixing modes for panels. There is some evidence that mixed-mode surveys can improve the representativeness of a sample. For instance, web only approaches exclude important and often underrepresented segments of the population. However, web surveys supplemented by a subsequent mail questionnaire produced results similar to those of mail-only surveys, but at an increased monetary cost to the researcher. There are also some additional concerns researchers must consider in using a mixed-mode approach. For instance, mixing modes can lead to differences in substantive responses due to one mode favoring a certain socio-demographic over another mode, e.g., RDD attracting older homeowners whereas web-only surveys are more apt to obtain responses from younger people. Another source of differences stems from mode of presentation with visual self-administered surveys often obtaining different answers than aurally administered questionnaires. In this research, we explore these issues with data from the NORC Amerispeak Panel, which recruits web panelist using a mail and telephone recruitment but subsequently follows up with a sample of non-respondents using in-person, face to face interview techniques. The results from this work focus on the demographic variability and changes in substantive responses relative to national benchmarks in order to understand the balance between potential mode effects and a representative sample.

Effects of Different Responsive Web Frameworks on Survey Response Quality Across Desktop and Mobile Devices
Alerk Amin, *RAND Corporation*

Many respondents complete web surveys using smartphones, and this number increases every year. There has been a great deal of research on the impact of device and screen size on response quality. Many surveys now use a responsive web design to provide a user-friendly interface to respondents, regardless of their device. There are many web frameworks that have been incorporated into surveys to implement responsive web design. Importantly, the choice of framework has large implications on the experience of respondents, from layout and design to functionality of input controls. Each framework has strengths and weaknesses on each device, and these lead can lead to varying respondent behavior and response quality. Two of the most popular frameworks are Twitter Bootstrap and jQuery Mobile. To our knowledge, previous research has not compared different frameworks.In this study we investigate the differences between Bootstrap and jQuery Mobile. To do this we use the RAND American Life Panel, a nationally representative probability based panel that participates in regular internet surveys. For one survey, we randomly assigned respondents to a version of the survey implemented using either Bootstrap or jQuery Mobile. The survey has varying question formats and controls, such as radio buttons, checkboxes, and text areas. These are rendered differently using the two frameworks. We analyze the impact of these two frameworks on response quality and respondent behavior across devices.

Tracking the Representativeness of an Online Panel Over Time
Thomas Klausch, *GESIS - Leibniz Institute for the Social Sciences*
Annette Scherpenzeel, *MEA*

One of the greatest threats to the accuracy of estimates from panel data is the occurrence of attrition of participants over the panel lifetime. Fortunately, panel recruitment interviews often provide rich sources of data that allow identifying the type of background characteristics affected most seriously by attrition. However, in practice the degree of attrition is seldom studied on these rich data sets, possibly because it is difficult to synthesize the degree of attrition across many auxiliary covariates. In this study, we address this shortcoming while studying attrition in the Dutch Liss online panel using the rich set of auxiliary variables available from the panel recruitment interviews. We estimated panel representativeness for each month of seven years of panel lifetime (2008-2015) using so-called representativeness (R-) indicators. The favorable feature of R-indicators is that they allow quantifying the degree of selectivity in a sample at a given point in time based on the information encoded in the auxiliary data. In doing so, we could include nearly forty variables in estimating representativeness, including respondents' health, economic status, personality, political attitudes, and survey attitudes, besides standard socio-demographics. The time-series of R-indicators that resulted from this project offers a unique view into the evolvement of panel representativeness and attrition. We found that attrition occurred in the Liss panel in two stages. First, there was an onset loss in representatives that occurred already at the first panel wave. Second, representativeness decreased across two and a half years of panel lifetime and subsequently stabilized. In our presentation, we show how the general loss can be split up into the contributions of the variable groups. We also evaluate the impact that a stratified and a random refreshment sample had on panel representativeness.

Increasing Response Rates: Cutting Edge Approaches in Design and Implementation

What's Your Number? Evaluating the Success of Telephone Number Acquisition Via Record Match, Mail Request, Web and In-person Follow-up Using an Address Based Sample

Lisa R. Carley-Baxter, *RTI International*
Julie C. Linville, *RTI International*
David Grant, *University of California, Los Angeles*
Matt Jans, *University of California, Los Angeles*
Royce Park, *University of California, Los Angeles*
Tara Becker, *University of California, Los Angeles*

Given the dramatic increase in cell-phone only households and the difficulty of geographically-targeting cell-phone samples, address-based sampling (ABS) is the best available approach to obtain a representative sample within a small, geographically-defined area (e.g., neighborhood or community). When the data collection mode is telephone, a telephone number must first be obtained for every sampled unit. This study evaluates the effectiveness of record matching, a mailed household information sheet, and a web response option for such a purpose. The California Health Interview Survey (CHIS) oversampled 14 communities as part of The California Endowment's (TCE) Building Healthy Communities (BHC) program. Initially 46% of sampled addresses were matched to a telephone number through two sampling vendors. The remaining addresses were mailed a household information sheet asking for phone numbers associated with the address and a best time to reach someone. Addresses for which the originally-matched telephone number was not useful were also mailed household information sheets. This sheet also included a link to a website where sampled households could complete the form electronically, with more than 5 times as many returns coming via mail compared to web. Sampled addresses received up to three mailings (initial letter, postcard, reminder letter) and a subset of non-respondents were eligible for in-person follow-up to collect the telephone number. We compare the rate of return for each community each contact, and between originally matched and originally unmatched cases. We also compare the screener and extended interview completion rates by timing of return (first mailing, postcard, second mailing) and mode of return (hardcopy, telephone, web, in-person).

Cell Phone Address Matching: A New Approach to Improving Survey Quality and Response Rates for Cell Phone Samples

Barbara M. Fernandez, *Abt SRBI*
Dennis Daly, *Abt SRBI*
Raphael Nishimura, *Abt SRBI*
Katherine A. James, *University of Colorado*
Yaqiang Li, *University of Colorado*
Arnold H. Levinson, *University of Colorado*

As the percentage of cell phone-only households continues to rise, declining response rates in cell phone frames is an increasing concern for survey researchers. It is, therefore, important to identify new methods to increase response rates and enhance the representativeness of cell phone samples. While advance letters are commonly used to increase response rates in landline RDD studies, this option has only recently become available for cell phones. The enhancement to cell phone sampling attaches

billing address to the cell phone number. Our study was conducted in 2015 with the University of Colorado, Denver, to determine the effectiveness of advance letters in increasing the response rates and improving the quality of the survey estimates in a recent dual frame (Random Digit Dial) RDD study, "The Attitudes and Behaviors Survey (TABS) of Colorado" (Colorado TABS). The research questions addressed in this study are: 1. How accurate is the matching of addresses to cell phone numbers? 2. Does sending an advanced letter to matched cell phone addresses improve the cell phone response rate? 3. Does sending an advanced letter to matched cell phone addresses improve the representativeness of the sample? The study includes an experiment to assess the impact of sending an advance letter to a random sample of cell phone cases with a known address compared with matched cell phone addresses not sent an advance letter, and cell phone numbers with no matched address. We will examine the differences between groups with respect to contact, cooperation, refusal, and response rates, and respondent sample composition including demographic characteristics and smoking status.

Influence of Multiple Factors on Response Rates
Bradford W. Chaney, *Westat*
Audrey Kindlon, *National Science Foundation*

Using built-in experiments and paradata, we examine the relative influence of multiple factors on response rates: prepaid incentives ($0, $2, and $5), variations in contact approaches (e.g., regular versus priority mail), web vs. paper surveys, quality of the contact data, relevance of the survey topic to respondents, questionnaire design (formatting and the ordering of questions), and length of the questionnaire. These factors are considered both alone and in combination, allowing us to examine: (1) which factors have the greatest impact, (2) the additive impact when all conditions are optimal, and (3) how some factors influence others (e.g., the quality of the contact data may influence the effect of providing financial incentives). Some results were surprising. For example, while most respondents preferred the web format over the mail format, and while more web responses were received than mail responses, the web format received higher response rates if people first received a paper questionnaire. Also, skip patterns appeared to lower the perceived relevance of the survey, resulting in lower response rates for mail surveys but not for web surveys. Based on these findings, we conducted a second round of data collection using a reserve sample, getting much higher response rates. This paper includes data from both rounds. The National Survey of Small Businesses, sponsored by the National Center for Science and Engineering Statistics within the National Science Foundation, was of 5,000 businesses with between one and nine employees, using a sample stratified by business organization type, industry, and number of employees. This audience is a difficult one to reach because of high mobility and sensitivity concerning the time required to complete the survey. Data collection was conducted by mail, with mail, email, and telephone followup.

Exploring Establishment Survey Incentives to Improve Response: The Impact of Class of Mail Delivery, Monetary Incentive and a Non-monetary Incentive on a Large Scale National Establishment Survey
James Dayton, *ICF International*
Robert Tortora, *ICF International*
Healey Whitsett, *Pew Charitable Trusts*

Conducting surveys of business organizations provides different challenges from conducting household surveys. In addition, there is much smaller body of research available investigating the best balance of various incentive types to encourage participation without biasing survey returns or

impacting item non-response. Incentives can be given to the establishment being surveyed or to the individual(s) actually completing the survey. Monetary incentives can be given before or after the survey is completed and mail notifications can provided via regular or certified mail. And unlike household surveys, an effective post-completion incentive may include survey results, especially those that benchmark the respondent organization against others in their cohort. This paper studies the impact on response rates and method implementation costs using various incentive and mail delivery options on a large scale establishment survey with 250 or fewer employees. These options are 1) class of mail delivery of a pre-notice letter, either 1st class or certified, 2) incentives, either no incentive, a prepaid cash incentive of $10 or a promise of a $20 incentive upon completion of the survey and 3) feedback on the results, either a promise of a report on the survey or no promise. We expect the margins of error for the control and treatments to range between +/- 4% and +/-8%/ The experimental treatments are embedded in the overall survey with 150 or more expected completed interviews for each condition. We anticipate that the prepaid incentive with either 1st class or certified will produce the highest response rates. To evaluate the effectiveness of these methods we will compare response rates, company profiles and item non-response based on the above conditions, paying particular attention the response impacts of pre-and post-incentives combined with two mail delivery options. Finally, we will attempt to measure the impact of a benchmark survey report on survey response.

Methodological Brief –
Measurement Issues and Measurement Error

Psychometric Scale Development for Arts Participation among Older Adults
LinChiat Chang, *www.linchiat.com*
Yongwei Yang, *Google*

The Health and Retirement Study (HRS) is a longitudinal panel survey designed to examine how older adults' changing health over time interacts with socioeconomic and psychological factors. The core HRS data is stocked with robust measures of physical health, cognitive functioning, and subjective well-being. The 2014 HRS included a supplementary arts module that has created an opportunity to study a broad array of relationships between arts participation and well-being of older adults. To facilitate such investigations, a valid and reliable scale is needed to assess values and perceptions relating to arts participation. To this end, this paper will provide a psychometric assessment of 8 potential items in the Arts and Culture module that were measured on the 5-point Likert (agree/disagree) response scale, to capture values and perceptions about the arts and participation in the arts. The sample of 1,500 respondents will be split into training vs. validation data sets. The training set will be subjected to pairwise correlations and exploratory factor analysis to map items onto coherent latent constructs of interest; competing factor solutions will then be fitted on the validation data set via confirmatory factor analysis to evaluate and identify the final recommended factor structure. Interim scales will also be assessed by Cronbach's coefficient alpha, item-total correlations, and tests of convergent and discriminant validity against other HRS indices where longstanding associations, or lack thereof, have been demonstrated in past research. Strength and limitations of agree-disagree scales will be discussed, with focus on the negatively-worded items to illustrate how the cognitive burden of having to process double negatives may be too much for less educated respondents. The final scale will be produced with proven psychometric properties to facilitate future investigations.

Changes in Satisfaction with Health Insurance Coverage Based on Question Order

Alisha Baines Simon, *Minnesota Department of Health*
Sarah L. Hagge, *Minnesota Department of Health*
Kathleen Thiede Call, *State Health Access Data Assistance Center*
Stefan Gildemeister, *Minnesota Department of Health*
Giovann Alarcon, *State Health Access Data Assistance Center*

The goal of health insurance is to provide financial protection from medical bills and facilitate access to health care services. In 2014, 80% of Minnesota adults were satisfied with the financial protection provided by their insurance. This question was asked in the 2015 Minnesota Health Access Survey (MNHA), a dual-frame RDD telephone survey, to allow for exploration of the relationship between satisfaction and use of and access to health care. However, the location in the survey was randomly assigned. Order effects in surveys occur when the response to a question depends on the order it appears in the survey. Further, a question's placement can alter the context of following questions by presenting information that may influence respondents' judgments. In this case, the question about financial protection offered by health insurance coverage was asked of all insured respondents either before or after a set of questions about health care utilization and access to health care. Preliminary results show lower levels of satisfaction when the question is asked after the assessment of health care utilization and access. The magnitude of order effects does not appear to change based on type of health insurance coverage, but for those who experience problems with medical bills or forgo health care due to cost, satisfaction is as much as 10 percentage points lower. To isolate the impact of order effects, multivariate analysis will account for sociodemographic factors which are linked to experiences of access barriers. Finally, using Item Response Theory (IRT) we attempt to detect whether item-level biases exist between the randomly assigned groups. These early results indicate that order effects may influence respondent judgments, especially in surveys which are trying to assess both experiences and satisfaction. For longer topic-based telephone surveys with both types of items, item placement must be carefully considered.

Implications of Response Device Type for Sensitive Web Surveys: Examining Data Quality and Respondent Characteristics in a Survey of College Students

Amanda C. Smith, *RTI International*
Ashley Richards, *RTI International*
Kimberly C. Peterson, *RTI International*

Mobile device use is increasingly common among young adults 18-29, 85% of whom report owning a smartphone (Smith, 2015). As mobile use has increased so has mobile survey response. While research on the impact of mobile responding is mixed, there is some suggestion that it can lead to data quality issues such as increased break-offs (Mavletova, 2013). These data quality concerns may be compounded when surveys contain sensitive questions. Previous research found that respondents trust mobile devices less with regard to data confidentiality, but saw mixed results when examining the differences in reporting rates of sensitive items between mobile and PC responders (Mavletova and Couper, 2013). However, this study primarily focused on attitudes toward deviant behavior (e.g., drug use) and randomly selected sample members from an online panel; findings may differ with other sensitive topics, such as sexual assault, and in young adult populations with higher rates of mobile use. In this paper we explore differences in reporting of sensitive items and item nonresponse rates for sensitive questions to better understand how survey estimates may differ by device type (computer,

smartphone, and tablet). We also examine the impact of device type on survey timing, as well as device type usage by respondent characteristics (e.g., sexual assault victimization rates). Data are from the Campus Climate Survey Validation Study Pilot Test, a Bureau of Justice Statistics and Office of Violence Against Women sponsored web survey of college students at nine U.S. institutions. Over 23,000 respondents completed the survey among a random sample of approximately 50,000 students. The survey included questions that measured rates of unwanted sexual contact and campus climate related to sexual harassment and assault. This paper concludes with a discussion on how survey practitioners can use results to inform the design of sensitive web surveys of college student populations.

Missing or Ignoring? A Case Study from the National Household Education Surveys Cognitive Tests 2015

Anna B. Sandoval Giron, *American Institutes for Research*
Caitlin Deal, *American Institutes for Research*

Survey methodologists continue to struggle with how to solve problems with skip patterns in mail questionnaires. Past experiments by Dillman et al. (1999) used college students to examine the effects of eight characteristics on error on errors of omission and errors of commission. Our interest lies in the impact of language and literacy on correctly following skip instructions. This paper aims to investigate the role of respondent choice and possible knowledge of what is termed the "form culture" on errors of commission using the National Household Education Surveys (NHES): 2015 cognitive testing. "Form culture" refers to the notion that individuals, starting from a young age, are required to fill out extensive amounts of paperwork (including testing, taxes, certifications, medical forms) and are expected to complete this paperwork without assistance. We contend that individuals most unfamiliar with the "form culture" will struggle the most with skip patterns and written questionnaires in general. In this case, those most unfamiliar with the "form culture" are those in the process of acculturation and adaptation to the U.S. model of forms and information. During the NHES: 2015 cognitive testing, we interviewed ninety seven Spanish dominant speakers and found that a forty five of the respondents missed at least one skip instruction in the questionnaire. During probing we explored the reasons why respondents missed the skips. Some respondents reported noticing the skip instructions but choosing to ignore these instructions. Upon further probing we found that respondents chose to continue answering questions because they feared that they would not provide all the necessary information. Results of this research provide a greater understanding of these patterns of behavior from respondents, improving our ability to design questionnaires that meet the needs of these and other respondents.

Demographic Characteristics of Respondents Who are More Likely to Have Problems Reporting Prior Day Events

Timothy Triplett, *Urban Institute*
Brian Tefft, *AAA Foundation For Traffic Safety*
Rob Santos, *Urban Institute*

Research has shown that surveys that require respondents to recall events can be subject to relatively high levels of measurement error. Recall error tends be less problematic for highly salient events and events that have recently occurred. However, there is less information on whether certain types of respondents are more or less prone to having problems with answering questions that involve recalling an event. This research uses data from the National Driving Exposure Study where people are asked to report the length of driving trips that they made yesterday. We analyze of over 14,000 reported driving trips from data collected from 4,744 respondents who reported having been the driver

for at least one driving trip on day prior to day they were interviewed (yesterday). For this analysis, we are concerned with two types of recall problems; 1) the inability of the respondent to provide an estimate of either the length or duration of a driving trip; 2) providing an estimate of miles driven that is inconsistent given the duration and purpose of the trip. We will be looking at the demographic characteristics of respondents who are more likely to have problems reporting their prior day driving behavior. Finally, we will provide some important considerations that survey practitioners should keep in mind when designing surveys that include recall questions.

Does Asking for Linkage Consent in the Beginning of the Questionnaire Affect Respondents' Answers?
Georg-Christoph Haas, *Institute for Employment Research (IAB)*
Stephanie Eckman, *RTI International*

Combining survey data with administrative data strengthens researchers' analyses, but we usually need the consent of each respondent to do so. Several studies have shown that the highest consent rates are achieved by asking the linkage consent question at the beginning of the questionnaire. However, asking this question at the beginning may impact respondents' answers throughout the interview. We hypothesize that asking for consent could affect responses in one of two contradictory ways. First, respondents might fear that their answers will be checked and thus expend more effort to give correct data (better respondents hypothesis). Second, respondents might think that their responses are no longer important and thus expend less effort (worse respondents hypothesis). We test these two hypotheses using a survey that asked half of the respondents for linkage consent at the beginning of the survey, and the other half at the end. Since the consenters in the two groups are not equivalent, we use the entropy balance technique to simulate a random distribution over both groups and calculate the effect of asking the consent question at the beginning on data quality, operationalized as measurement error (responses compared to administrative data) and item nonresponse. Preliminary results show that the linkage consent question does not affect response behavior, in the web or in CATI, which is good news for researchers who want to ask the linkage consent question in the beginning of the questionnaire, and also for methodologists who use linked data to investigate measurement error.

Survey Rating Scale Format Effects on Reliability: A Meta-analysis Summarizing the Impact of the Number of Response Categories
Michael Bosnjak, *GESIS - Leibniz Institute for the Social Sciences*
Natalja Menold, *GESIS - Leibniz Institute for the Social Sciences*

This meta-analysis is aimed at resolving one controversy around seemingly mixed findings on the relationship between rating scale format effects on measurement quality indicators which is prevalent for decades. Specifically, we synthesized the effects of the number of response categories in rating scales on reliability estimates, which had been published in 23 manuscripts between the years 1953 and 2013. Each of the studies identified reports diverse reliability coefficients for different numbers of categories. Overall, 189 reliability estimates along with the respective information about the number of scale categories and other variables potentially influencing reliability (e.g., scale polarity, category labeling scheme, topic) could have been extracted.Our findings show that the number of categories is positively, yet non-linearly related to scale reliability. The findings appear robust after controlling for other possible determinants of reliability. We will discuss these findings for development of survey items, and will sketch some avenues for future research.

The Impact of Interviewers on Data Quality and Participation: A Session in Honor of Charles F. Cannell

Panel organizer: Nancy Mathiowetz, *Independent Consultant*

Even in 2016, interviewers still matter! Despite declining use of interviewers, many surveys still engage the services of interviewers, who serve to both encourage participation and to assist respondents in the process of understanding the question and formulating responses that meet the goals of those questions. The Charles Cannell Fund in Survey Methodology was established by students, colleagues and friends of Charlie to honor him as a mentor and to further research and training on the interviewer-respondent interaction and its effects on the validity and quality of survey data. Since 2004, the fund has been supporting young investigators whose research focuses on the effects of interviewers, whether those effects are evidenced in rates of participation, the quality of data responses, or the willingness of respondent to divulge highly personal information. This session highlights but a few of the findings from research supported by the Cannell Fund, including the effects of rapport, familiarity, interaction, as well as interviewer dress and perceptions on survey data quality and participation.

Interviewing Practices, Conversational Practices, and Rapport: Responsiveness and Engagement in the Standardized Survey Interview

Dana Garbarski, *Loyola University Chicago*
Nora Cate Schaeffer, *University of Wisconsin-Madison*
Jennifer Dykema, *University of Wisconsin-Madison*
Ellen Dinsmore, *University of Wisconsin-Madison*
Bo Hee Min, *University of Wisconsin-Madison*

"Rapport" has been used to refer to a range of positive psychological features of an interaction -- including a situated sense of connection or affiliation between interactional partners, comfort, willingness to disclose or share sensitive information, motivation to please, or empathy. Rapport could potentially benefit survey participation and response quality by increasing respondents' motivation to participate, disclose, or provide accurate information. Rapport could also harm data quality if motivation to ingratiate or affiliate caused respondents to suppress undesirable information. Some previous research suggests that motives elicited when rapport is high conflict with the goals of standardized interviewing. We examine rapport as an interactional phenomenon, attending to both the content and structure of talk. Questions about one's end-of-life planning and preferences are potentially sensitive and interactionally delicate for both interviewers and respondents, creating a unique opportunity to study the development and maintenance of interactional rapport. Using questions about end-of-life planning in the 2003-2005 wave of the Wisconsin Longitudinal Study, we observe that rapport consists of behaviors that can be characterized as dimensions of responsiveness by interviewers and engagement by respondents. We identify and describe types of responsiveness and engagement in selected question-answer sequences. Our analysis suggests that responsive and engaged behaviors vary with respect to the goals of standardization—some conflict with these goals, while others complement them. We then devise a coding scheme of the identified types of responsiveness and engagement. We examine the analytic potential of these codes using the criterion of future study participation, as rapport developed during the interview may affect respondents' decisions to participate in subsequent data collection efforts. Using a case-control design, we examine whether interviewers' responsiveness and respondents' engagement measured during the 2004 end-of-life module predict respondents' likelihood to participate in a subsequent request to provide a saliva sample.

Revisiting the Interviewer-respondent Interaction in the Panel Study of Income Dynamics' Event History Calendar

Jamie Griffin, *University of Michigan*
Hanan Hashem, *University of Michigan*

Since 2003, the Panel Study of Income Dynamics (PSID) has utilized an interviewer-administered computerized event history calendar (EHC) to collect information about respondents' past-two-year residence and employment. A notable feature of the EHC is the conversational interviewing approach employed by interviewers. Previous research demonstrated that certain verbal behaviors facilitated by the conversational interviewing approach were associated with improved data quality compared to a standardized interviewing approach. Following the initial validation study, there have been no efforts to evaluate the performance of the PSID's EHC as actually implemented in production. Because the design of the EHC has changed over time, the extent to which the method continues to be implemented as intended remains an open question. Additionally, as the study's principal investigators consider translating the interviewer-administered EHC to web self-administration, it is particularly important to understand how this method is currently being implemented in CATI in order to maintain consistency in translation to web. Using recordings of the interviewer-respondent interaction from the 2013 PSID main interview, the current study has three aims. First, we document the prevalence of verbal behaviors, with a specific focus on those previously demonstrated to be related to data quality. Second, we (a) document the prevalence of respondent uncertainty about the timing of various residential changes and employment-related events (e.g., moves, job changes), (b) document how often each unit of time (e.g., exact date, best estimate of date, third of month) is recorded as the final response, and (c) examine whether respondent uncertainty is associated with the eventual response. Third, we (a) document how interviewers react to respondent uncertainty about the timing of residential and employment-related events and (b) examine whether interviewer reactions are associated with the eventual response.

Interviewer Effects in Panel Surveys: The Role of Panel Experience and Interviewer-respondent Familiarity

Simon Kühne, *German Institute for Economic Research (DIW Berlin)*

A large body of cross-sectional research documents that interviewers affect respondent behavior and responses during personal interviews. These "interviewer effects" reduce overall survey data quality by introducing variance and bias into estimates. Yet, despite the growing popularity of longitudinal studies, interviewer effects in panel surveys are still largely unexplored. This paper focuses on socially desirable responding, which has been shown to be responsive to the characteristics of respondents, interviewers, as well as their interaction. Generally, the paper argues that socially desirable responding reduces with every wave of a longitudinal survey. Three processes may constitute this effect: First, the cumulative survey experience of respondents may foster trust in the study and thus to reduce socially desirable responding. Second, I hypothesize that the cumulative survey (work) experience of interviewers enables them to elicit more accurate answers on sensitive topics and thus, again, to reduce socially desirable responding. Finally, I test whether or not the growing familiarity between individual interviewer and respondent in panel surveys increases the respondent's trust and consequently reduces socially desirable responding. The analysis is based on data emanating from the German Socio-Economic Panel (SOEP) study. The longitudinal study, which covers more than 30 annual panel waves, is unique in providing longitudinal information not only for respondents, but also for interviewers.

Interviewer Religiosity and Polling in Transitional Tunisia

Lindsay J. Benstead, *Portland State University*
Dhafer Malouche, *University of Carthage, Tunisia*

How and why do interviewer traits, including religious dress, affect polls in the Middle East and North Africa (MENA)? One potential cause of results that vary widely from one firm to another, and from actual election results, may be response and non-response bias stemming observable interviewer traits. Respondents infer interviewers' political affiliations, including whether they support a secular or Islamist party, based on interviewers' dress or known affiliation of the firm conducting the research. Using three nationally-representative surveys of 1,200 Tunisians conducted in 2012 and 2014 and 3,600 Tunisians in 2015, this paper assesses the link between interviewers' observable traits and responses to questions about past and future vote choice. Interviewer religious dress increases the likelihood of stating one voted in the past for the Islamist Ennahda party, or plans to in the next elections. Effects are small for religious female and large for religious-appearing males. Implications for reducing Total Survey Error in polls in the MENA region are considered.

Explaining Interviewer Effects on Nonresponse Across Two Face-to-Face Surveys

Julie M. Korbmacher, *Max Planck Institute for Social Law and Social Policy*
Ulrich Krieger, *University of Mannheim*

In this presentation we will discuss the results from a comparative analysis of interviewer effects on contact and cooperation in two face-to-face surveys: The German Internet Panel (GIP), and the German implementation of the Survey of Health, Aging and Retirement in Europe (SHARE). Survey interviewers have been found to have different success rates when contacting respondents and gaining their cooperation leading to interviewer effects on nonresponse in face-to-face surveys (Blom, de Leeuw, & Hox, 2011). Nonresponse rates also vary between surveys due to different field procedures, samples or topics. Using data from the GIP recruitment interviews in 2012 and 2014 and the SHARE Germany 2014 refreshment sample we can jointly analyze interviewer and study effects on nonresponse. Both studies used a harmonized interviewer survey capturing values, opinions and attitudes of interviewers prior to field start. In addition, both surveys were fielded by the same survey agency reducing organizational effects on nonresponse patterns. This unique setting allows to compare the effects of interviewers' characteristics on contact and cooperation across the two surveys and disentangle study and interviewer effects on nonresponse. Preliminary results indicate interviewer effects on cooperation to be influential and to vary between the surveys examined.

The Relationships between Interviewer and Respondent Speech, Rapport Evaluation and Data Quality

Hanyu Sun, *Westat*

Building a good relationship with respondents, or the establishment of rapport, is frequently mentioned as important in interviewer training materials (e.g., Adams, 1958), and it is often speculated that it affects the quality of data obtained in survey interviews. Behavior coding has been used to measure rapport-related conversational behaviors, such as interviewer/respondent digression, interviewer/respondent laughter, and improper or directive probing (e.g., Belli, Lee, Stanford, & Chou, 2004; and Belli, Bilgen, & Baghal, 2013). The findings are mixed—rapport-related conversational behaviors sometimes improve data quality (e.g., Belli et al., 2013); at other times they impair data

quality (e.g., Belli et al, 2004). These prior studies often associate rapport with speech behaviors that deviated from the protocol for standardized interviewing. However, they did not verify that these speech behaviors are indeed measures of rapport instead of some relevant constructs such as liking or agreement. In addition, rapport evaluation from the conversational partners—interviewer and respondent—was not collected in previous research, which makes the investigation of the relationship between interviewer speech and rapport impossible. Although standardized interviewing does not provide as much opportunity as conversational interviewing to establish rapport, interviewers may still be able to develop and maintain a good relationship with the respondent without violating the protocol for standardized interviewing (e.g., Lavin and Maynard, 2001). This study explores the relationships between interviewer and respondent speech, rapport evaluation, and data quality (e.g., sensitive disclosures and item nonresponse). Behavior coding of both interviewer and respondent verbal behaviors during the questionnaire administration in a laboratory experiment— the respondent and interviewer in each interview both rated rapport—will be used to examine whether interviewers rated high in rapport are conversationally different than interviewers rated low in rapport as well as whether the presence of certain speech behaviors (e.g. quasi-laughter) are associated with higher data quality.

Rise of the Machines:
Machine Learning in Social Research

Employing Machine Learning Approaches in Social Scientific Analyses
Arne Bethmann, *Institute for Employment Research (IAB)*
Jonas F. Beste, *Institute for Employment Research (IAB)*

With the increasing availability of both new large scale data sources as well as computational power, specific algorithms for the analysis of these data have seen widespread use and many refinements. Certain major developments in this area stem from computer science and computational statistics and are known as machine / statistical learning approaches. While these algorithms are widely used for predictive purposes their application to statistical inference is rather scarce. In comparison to traditional modelling approaches in the social sciences, especially (generalized) regression models, machine learning algorithms often allow for more flexible modelling strategies. Depending on the specific algorithm there are additional desirable features, for example automatic variable selection or implicit specification of interaction effects. Research questions in the social sciences usually aim at population inference for effects of certain variables of interest rather than prediction. Therefore a way to draw inference from the resulting machine learning models has to be developed if they are to be applied in social scientific research. We propose an approach adapting the idea of average marginal effects in order to draw inference from a variety of classical machine learning algorithms like e. g. Decision Trees, k-Nearest-Neighbors, Naïve Bayes. We will provide results from a simulation study focusing on bias and variance of the different estimators In addition to new modelling features machine learning algorithms are in many cases better suited for large datasets than traditional modelling approaches. The Institute for Employment Research (IAB, Germany) is a large producer of social scientific data. Most notably they provide vast amounts of data on complete employment biographies for labor market related research. We will use examples from this area of research to compare the different modelling approaches for substantial social scientific research.

Using Big Census Data to Better Understand a Large Community Well-being Study: More than Geography Divides Us

Donald P. Levy, *Siena College Research Institute*
Meghann Crawford, *Siena College Research Institute*

Based upon 16,000 interviews in one state, this study of community well-being offers an extensive look at the health, satisfaction, neighborhood perception, access to medical care, economic strength, and personal safety and includes robust samples by not only sub-geographies but also by age, race/ethnicity education and income. Separately using three U.S. Census variables – median family income, poverty percentage and population density by town – the towns of this state are grouped via K-means cluster analysis and discriminant analysis into five distinct groups. These socio-economic/culturally variant groups are then used as the lens through which an analysis is done of the community well-being data. Virtually every aspect of life varies substantially by income. But, it is more than just how much money an individual household earns but rather the community in which one lives as determined by not only income but also population density and the presence of poverty that fashions the state into five distinct clusters – the Five Connecticuts. The Five worlds while geographically interspersed are clearly divided. For example, in 'Wealthy' Connecticut over ninety percent of residents rate the job done by the police as excellent or good while nearly half in the 'Urban Core' rate the police's efforts as no better than fair or poor. The five groups are titled: Wealthy, Suburban, Rural, Urban Periphery and Urban Core. This paper describes how the towns were analytically placed into clusters at three successive time points: 1990, 2000 and 2010. The data from this 2015 community well-being study is then analyzed across these five different groups. This paper demonstrates the successful analytic collaboration of Big Data (Census) and public opinion research, in this case, amplifying the available insight offered by the public opinion data as well as potentially motivating implementations by policymakers, service providers and citizens.

Using Machine Learning to Infer Demographics for Respondents

Noble Kuriakose, *SurveyMonkey*
Tommy Nguyen, *SurveyMonkey*

In this study, we assess machine learning classification techniques. We particularly consider the effectiveness of algorithms to accurately classify respondents into demographic groups including age, gender, and education. We use a feature rich data source-- SurveyMonkey Intelligence. Data are collected directly from panelists' mobile devices and contains a trove of information like length of phone usage, website and app usage, time of day and weekly patterns. We use training data from over 20,000 unique panelists who answered a web survey to provide their demographic information and shared their mobile usage data to test different models. We consider the breadth of algorithms available to tackle classifications problems like this one, report on the tradeoffs of using tree based approaches as opposed to clustering algorithms, and provide guidelines to successfully use a high-dimensional dataset to make inferences about respondents' demographics. We suggest some steps forward for researchers who have behavioral data for all respondents but see high levels of missingness on demographic data.

Dissonance and Harmony: Exploring How Data Science Helped Solve a Complex Social Science Problem

Michael L. Jugovich, *NORC at the University of Chicago*
Emily White, *NORC at the University of Chicago*

In 2010 NORC began working with the US Department of Justice Bureau of Justice Statistics (BJS) on an expert system capable of solving a complex data harmonization problem. The challenge was to take semi-structured criminal history record information and consolidate the text fields into a series of uniform and standardized variables that enable recidivism research accounting for variations in state and federal jurisdictions. To solve this challenge NORC, supported by BJS, put together a team of criminal justice researchers and data scientists to develop an expert system capable of ontological classification by leveraging a sophisticated inference engine on an extensive knowledge base. After three years of development, the system completed the conversion of the first cohort of prisoners released in 2005, and since then, has processed multiple additional cohorts. In addition to processing additional cohorts, the system is being continually refined and improved, including the development of new natural language processing and machine learning models that support the core capabilities of the expert system. This paper will begin by discussing a brief history of the project including the motivations for the research and examples of how the datasets produced by the system have helped to enable multistate recidivism analyses. After this introduction, the paper will provide insight into how NORC blended social science research with data science to develop this expert system. By discussing the overt complexity of the research challenges and the innovative methods used to surmount these challenges, we intend to illustrate how data science is not only useful for social science research it is necessary for its long term prosperity.

Machine Learning Our Way to Happiness

Pablo Diego Rosell, *The Gallup Organization*

Research on happiness has exploded in the last 20 years, with significant contributions pouring in from the fields of Psychology, Economics, Sociology, Political Science and the Life Sciences. Researchers studying subjective well-being (SWB), the academic term for 'happiness', have proposed a multitude of statistical models attempting to identify the determinants of happiness. Yet most models suffer from relatively low explanatory power: The strongest SWB models hardly manage an R2 of .40, with most typical estimates varying between .03 and .15 (Senik, 2011), and generally lower explanatory power in Asia, Africa and Latin America (Helliwell et al., 2009). Assuming that we can control for measurement error, there are two major reasons left for such low explanatory power. The first one is omitted variable bias. Since different things make different people happy (Kroll, 2010), failing to incorporate a relevant determinant, even if it is not universally relevant, would bias the overall model. The second one is incorrect functional form specification. Failing to account for non-linear effects and interactions between predictors will lead to biased estimates of predictor coefficients and lower overall model fit. In this study we test the explanatory power of two different approaches: 1) a theory-driven approach incorporating variables and functional forms identified in the literature, and 2) a data-driven approach, based on machine learning algorithms that eschew theoretical assumptions. We test both approaches in the largest existing dataset on SWB, the Gallup Daily Poll. The study will identify the main determinants of SWB in the US, the relative importance of each determinant, and will test the hypothesis that automated algorithms can improve the predictive power of SWB models by "learning" the correct model specification directly from the data.

Mo' Money Mo' Respondents:
Using Incentives to Increase Survey Participation

Increasing Incentive Salience: Effects of a Pre-notification Letter on Nonresponse Follow-up and Nonresponse Bias Potential

Julie C. Linville, *RTI International*
Matt Jans, *University of California, Los Angeles*
David Grant, *University of California, Los Angeles*

Pre-notification letters are generally sent before the first household contact in telephone surveys, but could also be effective before nonresponse follow-up. If essential survey conditions are changed at this stage (e.g., incentive increase), increasing salience of that modification should increase contact, cooperation and response rates, and potentially reduce nonresponse bias (Groves, Singer, and Corning, 2000; Groves, Couper, Presser, and Singer, 2006). The California Health Interview Survey's Building Healthy Communities study is a telephone survey using an address-based sample of about 28,500 addresses in 14 geographically-, linguistically- and demographically-diverse communities across California. Households were called after providing a phone number on a mailed household information form or after their address was matched to a phone number. At telephone contact, a screener was conducted to select household members to interview (up to one adult, child, and adolescent). A sub-sample of telephone nonrespondents was selected for additional nonresponse follow-up by telephone. A random half of the nonresponse follow-up group received a pre-notification letter informing them of an increased incentive ($40, up from $20) and our upcoming call. Half did not receive a letter, but were called and offered the same increased incentive. We hypothesize that the mailing before nonresponse follow-up calling increases incentive salience, thus reducing follow-up effort and increasing response rates. The increased incentives overall, and specifically in the increased salience (i.e., letter) group, may also reduce nonresponse bias (as measured by the difference between screener respondents and nonrespondents on key health and demographic measures, and by comparing the change in key health and demographic estimates based between pre-follow-up cases and the complete data). We use call record paradata to control for differences in nonresponse follow-up effort, which could also increase incentive salience. Results will be assessed by community to assess geographic variability in the incentive and salience effects.

Perceptions of the Two Dollar Bill and Value as an Incentive

Jenny Marlar, *The Gallup Organization*
Manas Chattopadhyay, *The Gallup Organization*
Steve Ander, *The Gallup Organization*
Kirti N. Kanitkar, *The Gallup Organization*
Rob Andrews, *NOAA*
Ryan Kitts-Jensen, *NOAA*
John Foster, *NOAA*

Prepaid incentives are an effective way to improve response rates in mail surveys, and small token incentives of $5 or less have been shown to significantly improve response rates while minimizing survey costs. Because of this, nominal incentives, including a $2 incentive, are often a popular choice for researchers. However, little research has been done on differences between currency types, and findings are very limited and mixed on the effectiveness of a $2 bill versus two $1 bills. A $2 bill is novel and perceived by some as lucky or rare or of greater value than $2 and may seem more like a gift than

$1 bills. However, some people are unaware that a $2 bill is legal tender and some retailers even refuse to accept the bill. Alternatively, $1 bills may seem easier to spend or psychologically feel like "more" because the recipient receives two bills. However, $1 bills may be more expensive for the researcher because it requires inserting two bills into the survey packet rather than one. All of these factors may influence which denomination is preferable, but it is unclear which may have the advantage, and benefits may differ by demographics. A mail survey conducted by Gallup on behalf of NOAA fisheries tested the difference between two $2 incentives – two $1 bills versus a $2 bill. Respondents were randomly assigned to receive a $2 bill or two $1 bills in their initial survey packet. Non-response follow-ups did not include additional incentives. The findings will explore overall differences in response rates, response rate differences by geography, demographics, substantive answers, timeliness of response, and differences before and after reminder materials are sent. Perceptions of $2 bills will also be discussed, using data from a Gallup Panel survey. The cost implications of denomination choice will also be considered.

Looking Beyond Obligation: The Social Psychology of Prepaid Survey Incentives
Thomas M. Guterbock, *University of Virginia*
Shigehiro Oishi, *University of Virginia*
Casey M. Eggleston, *U.S. Census Bureau*

The positive effect of token, prepaid monetary incentives on response rates is well known but poorly understood. A combination of laboratory and field-based survey experiments helps uncover how prepaid incentives affect the experience of the survey request. Our lab experiment allowed us to measure key variables at the time of the survey request, including data from non-responders. Our field experiment allowed us to explore possible cultural differences in the degree to which prepaid incentives incur a sense of obligation. In the lab, 215 students participating in social psychology experiments were invited to respond to a survey invitation that would be sent later, with half receiving a $2 advance incentive and half receiving a promise of $2 contingent on completion. Immediately thereafter, students completed a questionnaire that asked them about ten current mood states, including their sense of obligation. As expected, students who received the prepaid incentive were more likely to respond and felt significantly more obligated. Among survey completers, those in the prepaid treatment retrospectively reported higher obligation. In the field experiment, an ethnically targeted sample of 2250 households was sent a mailed request to complete an online survey, with half getting a $2 prepaid incentive and half getting none. The questionnaire included items about obligation, guilt, survey legitimacy, importance, confidentiality, and appreciation. Prepaid respondents returned more questionnaires and had higher levels of obligation and hypothetical guilt. But the incentive also affected perceptions of importance, worthwhile effort, and feeling appreciated. These pilot results substantiate our argument, following Dillman et al. (2014), that the prepaid incentive positively affects a range of perceptions that are key to the Tailored Design Method: legitimacy, trust, and perceived costs of the survey task. In addition, Asian respondents had higher levels of felt obligation and hypothetical guilt, suggesting that incentive effects are culturally and socially conditioned.

Promoting Participation in Web Surveys
Andrew L. Hupp, *University of Michigan*
Will Chan, *University of Michigan*

Web surveys generally attain lower response rates than other modes of data collection, and response rates seem to be declining over time. Web surveys are also at risk of other forms of nonresponse, such as breakoffs, which happen less frequently in other modes. These breakoffs

commonly occur on informed consent screens, often required by IRBs, before they have a chance to get to the survey content. There are many methods used (prenotification, incentives, etc.) to try to increase participation and reduce breakoffs. This paper focuses on two experiments designed to assess the impact in increasing participation and reducing breakoffs. These two experiments were implemented on the Sustainability Cultural Indicators Program (SCIP) annual survey at the University of Michigan. The SCIP program monitors the U-M's progress in moving toward a culture of sustainability. Progress is determined by tracking a set of cultural indicators derived from responses to a yearly campus-wide sustainability questionnaire. The sample design includes a rotating panel design for undergraduate students and a yearly independent cross-section of faculty, staff and students. The first experiment examines if "celebrity endorsement" increases participation. As part of the final email reminder respondents were randomized to receive a regular reminder with a link to the survey or a reminder that also contained a link to a video of a head coach from the U-M Department of Athletics encouraging non-respondents to participate. The second experiment looks at the design of the informed consent screen. One group was presented a screen appearing as a traditional informed consent form. The other group was presented a screen with the most important items visible and the rest of the information available under a series of accordion menus.

Improving Methods to Survey Hispanics

Language Negotiation on a Bilingual (English/Spanish) Telephone Survey in the U.S.

Casey Langer Tesfaye, *Nielsen*
Carlos Jarava, *Nielsen*

The negotiation of language in telephone interviews where there is potential for a bilingual respondent is often debated. Should telephone interviews begin in Spanish if data indicates the possibility of reaching a Hispanic respondent or should they always begin in English? Each strategy has its advantages and disadvantages. Starting in Spanish can help Spanish speakers feel comfortable more quickly but may drive some English speakers away, and starting in English can have a wider appeal but draw less Spanish respondents. However, the language negotiation process is more nuanced and less conscious than these debates would have us believe. Not only is the first conversational turn taken by the interviewee and not the interviewer, language choice is ultimately a negotiation that involves both speakers. For this study, we examine a corpus of over 300 phone calls from a bilingual study to behavior code the mechanics of the language negotiation process. For each call the language of the first and subsequent conversational turns are coded as well as whether there was any shift in language during the conversation. We then take observational notes about interactions that involve more nuanced negotiation or any shift in language. In this paper, we will first present the overall findings about language choice and language shifting and then explore a selection of phone calls in greater depth to discuss the effect of language negotiation strategies on the progress and outcome of the call. The goal of this research is to develop data-driven, practical recommendations regarding language negotiation in telephone interviews.

What to Say after "Hola"–Surveying U.S. Hispanics "in-Culture"

Carlos Jarava, *Nielsen*
Victoria J. Hoverman, *Nielsen*
Courtney Mooney, *Nielsen*

The Census Bureau estimates that 54 million people in the United States (17% of the population) are Hispanic. Of those, 64% were born in the U.S., and 36% were born outside the U.S. If Latinos in the United States were a country, it would be the second largest Spanish-speaking country in the world. Utilizing a universal approach to research, without considering the cultural aspects of this group, will lead to inadequate representation of the complete U.S. population. In a series of qualitative studies, we explored culturally appropriate methods for surveying Hispanic populations. Three themes emerged from this research: (1) the importance of establishing trust with the respondent, (2) an understanding of the cultural differences between Hispanics and non-Hispanics, and (3) effective communications throughout the survey process. Establishing trust is critical, especially in a population in which there may by heightened concerns about privacy and confidentiality. In addition, assuming a universal value system is insufficient when trying to reach and survey Hispanics. It is necessary to understand key cultural differences within Hispanics and Hispanics compared to other ethnic groups. An understanding of what motivates Hispanics to participate in research can inform the development of effective methods for reaching this population. Further, when designing research materials and survey instruments, researchers should take a bicultural approach that effectively communicates with Hispanics; for example, by creating a unique voice in Spanish that will yield the same intended results, rather than utilizing literal English to Spanish translations. This presentation will explore lessons learned from in-depth interviews and focus groups with Hispanics on their attitudes towards survey research. Specifically, how can researchers go beyond "Hello" or "Hola" in an interview and speak to Hispanics "in-culture"?

The Impact Of Survey Language Administration And Socioeconomic Status On Skip Error Rates

Mahi W. Megra, *American Institutes for Research*
Anna B. Sandoval Girón, *American Institutes for Research*

Obtaining accurate data from a paper questionnaire depends heavily on respondents following skip instructions correctly. Wang and Sha (2010) found that there were significantly more skip errors for survey forms completed in Spanish than for forms completed in English, and that the increase in the error rate for Spanish-language forms was higher among respondents who had not completed high school. Nonetheless researchers have found mixed results when examining the role of education of respondents and the impact it may have on rates of missing data (McBride and Cantor 2010). Thus, it is important to assess what other factors might exacerbate or mitigate the impact of survey language on skip error rates. In the present analysis, we seek to investigate the role of socioeconomic status (SES) on skip error rates using data from the 2012 National Household Education Surveys (NHES: 2012). The NHES: 2012 was a primarily paper based two-phase survey. During the first phase a nationally representative sample of households were selected to fill out a screener listing all members under the age of 20. The second phase entailed administering either of two surveys based on the age of the selected child. Households in the Hispanic sampling stratum (constructed based on Census tract with more than 40% Hispanic persons) received both English and Spanish screeners. Households that returned a Spanish screener received a Spanish topical questionnaire in the second phase. Limiting the analysis to those households that returned the screener and topical questionnaires, a composite

measure of SES with indictors such as household income, and education level will be constructed. Skip error rate will be analyzed separately by questionnaire language and SES to assess whether respondent SES influences the magnitude of any language effect. Findings will help improve questionnaire design that responds to the needs of various target populations.

Challenges of Surveying Hispanics: The Importance of Cellphones

Anna Brown, *Pew Research Center*
Eileen Patten, *Pew Research Center*
Mark Hugo Lopez, *Pew Research Center*

Hispanic adults (59%) are more likely than non-Hispanic white (43%) and non-Hispanic black (48%) adults to live in a household with a cellphone but no landline telephone. That has implications for how to reach a representative sample of Hispanics by telephone, yet Hispanics are typically sampled on cellphones at the same rate as non-Hispanics. Using data from Pew Research Center's 2015 National Survey of Latinos, this study compares the characteristics of Hispanic adults who have a cellphone but no landline with those who have a landline. Cellphone-only Hispanics are more likely to be Spanish dominant, less likely to be bilingual, and are younger than their counterparts with landlines. This study presents a case for conducting a higher share of Hispanic interviews on cellphones compared with non-Hispanics.

Public Opinion and Political Engagement Among Minority Populations within the American Electorate

Panel organizer: Samara Klar, *University of Arizona*

The AAPOR Diversity Committee is proud to present this special panel, featuring some of country's most preeminent experts on issues relating to minority opinion and engagement. These six scholars will discuss their research, which addresses immigrant communities, racial and ethnic minorities, and gender diversity. This work has implications for how pollsters and academics can investigate and understand an increasingly diverse American population.

Capturing Complexity of Diverse Populations and Key Concepts and Relationships

John Garcia, *University of Michigan*

Surveys, of all types, can represent attempts to "capture" important "aggregation" of people and solicit ideas, perspectives, knowledge, attitudes and behaviors that affect one's everyday experiences. While constraints of time affects efforts of parsimony and efficiency, continual challenges remain. Two critical ones are the design of surveys to integrate the diversity of populations(with sufficient clusters of "characteristics"), and the inclusion of the breadth and depth of items that enable investigators to analyze key possible relationships that impact people's everyday experiences. An ancillary aspect of these challenges is the manner and mode of this type of data collection (i.e. language, cultural sensitivity/appropriateness, multiple modes of data collection, capturing multi-dimensionality of complex concerns). As a result this presentation will try to address these challenges in terms of consequences of not facing and addressing these challenges and some possible alternative approaches to address these challenges. This discussion will be cognizant of basic principles central to the conduct of survey research, as well as the practical constraints and considerations. If there is any expectation as

result of this presentation is a more converted discussion to acknowledge and explore alternative ways to use survey research to enhance the analytical power of survey research.

The Color of Our Skin & the Content of Our Politics: The Effects of Skin Tone on Policy Preferences Among African Americans
Vincent L. Hutchings, *University of Michigan*

Social scientists have established for some time that darker-skinned African Americanshave lower incomes, encounter more discrimination in the workplace, and have poorerhealth outcomes than lighter-skinned Blacks. These findings persist into the 21st century,yet public opinion scholars have found few attitudinal differences among Blacks thatreflect these disparate outcomes. That is, previous work has failed to uncover consistentcorrelations between skin tone and racial group identification or policy preferences. Werevisit this question by relying on the discipline's most comprehensive political survey:the 2012 American National Election Study (ANES). The 2012 ANES includesinterviewer assessed skin-tone measurements along with an oversample of Blackrespondents (N=511). We find significant skin tone related differences among Blacks oneconomic redistribution, explicitly racial policies, and endorsement of racial stereotypes.We also find in a supplemental experiment that skin tone based identities can be mademore relevant with respect to racial group identity.

Women as Contributors: An Exploratory Analysis
Kristin Kanthak, *University of Pittsburgh*

It is a truism in politics that women make campaign contributions at a much lower rate than do men. Only about 25 percent of campaign contributions come from women, and women are more likely than men to give to women candidates. This context provides an explanation for why pundits and academics were shocked when Hillary Clinton's presidential campaign released its July 2015 quarterly campaign contribution report showing that fully 66 percent of Hillary Clinton's contributions from April to July 2015 came from women. This paper is part of a larger project to analyze what this astounding difference might tell us about how women and men differ in their contribution patterns. In this first step, I seek to gather and analyze data on these contributors to better understand extant contribution patterns of Hillary Clinton's women donors in an attempt to explain why contribution patterns in this election do not mirror those we have seen before. What accounts for the large contributions from women? Are these new donors, making contributions for the first time? Or are they serial donors who are making larger contributions than they had before? This is the first step in a larger project that will continue over the next several elections.

Intergroup Relations in Urban Cities: 50 Years After Watts
Valerie Martinez-Ebers, *University of North Texas*
Brian Calfano

August 11, 2015 was the 50th Anniversary of the Los Angeles (LA) Watts Uprising/Riots. To commemorate the anniversary, a series of media, educational and community-based programs were held throughout LA county, that highlight issues relating to social justice, community policing, and racial-ethnic group relations. The assumption behind many of these programs is that heightening awareness of past challenges that the Watts turmoil typified (and the efforts made to overcome them) will strengthen inter-group relations and promote a shared or superordinate identity among the various subgroups of Angelinos. While some research supports this assumption, other studies have found increased

intergroup conflict or more complicated outcomes from similar activities. None of the existing studies are definitive due to data limitations. The anniversary programing provided a unique opportunity to address this controversy by creating the conditions of a natural experiment in which the prevailing community discussions about inter-group relations occured within the context of heightened awareness of both ingroup and outgroup distinctiveness. Against this backdrop of heightened group identity salience, with the sponsorship of the LA Human Relations Commission, a survey experiment with 1,300 respondents was conducted to measure the effects of different group identity cues in reducing (or not) inter-group conflict. The survey also included items gauging the effect of urban stressors and perceptions of local law enforcement. Taken together, these outcome measures will provide a holistic assessment of the state of intergroup relations in Los Angeles that can be used by to develop policy and programming recommendations for the Human Relations Commission, community partners, city council, and the mayor's office. The study findings could also be relevant for other communities with heightened intergroup tensions. This includes areas where recent community and law enforcement controversies have raged, such as Baltimore, Ferguson, and Charleston.

Is Demographic Data Objective? The Rise of Identity Politics and Measurement Using Self-reported Survey Items
Natalie Masuoka, *Tufts University*

Scholars who rely on survey data generally take demographic characteristics such as age, race, gender and education as "objective" indicators of a survey respondent's background that serve as useful controls to disaggregate a survey population. However, we are also aware of the fact that today, demographic characteristics are gathered through survey measures which allow the respondent to self-report his/her own sense of those characteristics. Using racial demographic data as a case study, this presentation will first review the different ways we have collected data on racial background and will highlight how our generally accepted trend to rely on the respondent's self-reported race reflects a distinct shift in our orientation towards race in today's political culture. Reflecting on these historical trends, I offer an argument that our demographic survey items are increasingly accounting for subjective personal identification, not necessarily objective demography as we traditionally define it. I highlight two interesting trends to consider: the rise of individuals who choose to self-identify as two or more races and the growing demands to recognize transgender identities.

Racial Symbols and Opinion Among the American Electorate
Tasha Philpot, *University of Texas at Austin*

Dr. Tasha Philpot is an Associate Professor of Government at the University of Texas at Austin. She is also affiliated with the Center for African and African American Studies and the Center for Women's and Gender Studies. Dr. Philpot's particular interests are in African-American Politics, Political Psychology, Public Opinion and Political Behavior, Political Communication, and Political Parties. In her book, "Race, Republicans, and the Return of the Party of Lincoln" (2007, University of Michigan Press), Dr. Philpot examines the circumstances under which political parties can use racial symbols to reshape their images among the electorate. She is also the co-editor of "African-American Political Psychology: Identity, Opinion, and Action in the Post-Civil Rights Era" (2010, with Ismail K. White, Palgrave Macmillan Press). Her work has been published in The American Journal of Political Science, Journal of Black Studies, PS: Political Science and Politics, Political Behavior, Public Opinion Quarterly, National Political Science Review, and the Journal of Politics. Dr. Philpot received her Ph.D. in Political Science from the University of Michigan.

Defining Data Science and Its Critical Place in Our World

Panel organizer: Michael W. Link, *Abt SRBI*

"Data Science" has many meanings depending on one's perspective, but it's growing impact and utility across the research world is undeniable. What does this mean for AAPOR members, our organizations, and those with (or for whom) we work? An international panel of experts explores a series of critical areas to better understand how data science is driving new directions in opinion, attitudinal & behavioral research across multiple sectors of our industry. Representing contract research, the private sector, academia and government, discussion focuses on several key questions: How do we define "data science"? What is the current and future impact of this type of research on various sectors? How is it changing research organization in terms of strategy, capabilities and the types of skill sets being sought in new hires? The implications of these and related questions will have a profound impact on the future of survey research and our broader industry.

View from the Commercial Sector
Annie Pettit, *Peanut Labs*

Big data creates a sense for fear and doubt for many market and social researchers who have little experience with it. This fear has been created by an abundance of myths related to data quality, scientific processes, domain knowledge and more. However, those same researchers often do not realize that their own skills and expertise lend themselves perfectly to owning this supposedly new domain. Indeed, they may not even realize that the work they've been doing for the last ten or twenty years may qualify to be called big data analytics. Consider this speaker to be a mythbuster!

View from Not-For-Profit Research
Craig Hill, *RTI International*

In approaching "Data Science," we typically use the definition proposed by the National Science Foundation in 2010: "Scientific and technological means of managing, analyzing, visualizing, and extracting useful information from large, diverse, distributed, and heterogeneous data sets." In practice, however, we need to separate the hype from the reality. Gartner's hype cycle (2015) suggests that "Big Data" and "data science" are both entering the "trough of disillusionment". Why? Because it's not as easy to gain knowledge/insight from massive amounts of data as people thought it would be. Why is that? Science/scientists are only now coming to grips with a host of issues presented by big data: data storage/storage costs, distributed data sets, access to data, useful lifetime issues, validation and curation. The "big data" landscape is crowded with players and new entities are created every day, but not all of these are adept at truly understanding the underlying data, the errors in those data, and how those errors can effect conclusions/"insights."

View from Government Statistics Organization
Lilli Japec, *Statistics Sweden*

      The explosion of data has facilitated the use of "big data' in place of traditional survey approaches for measuring some government statistics. The bar is high, however, and the challenges near endless in making this transition broadly. The Total Error perspective provides a useful framework for delineating and assessing these challenges. Additionally, as an organization in Europe, the privacy and data transfer laws and regulations are changing quickly and can be quite restrictive. Discussion will focus on how some of these challenges have been overcome in using large, non-survey databases to develop certain key statistical indicators for government use.

View from Academia
Frederick G. Conrad, *University of Michigan*

      Over the last 20+ years we have learned how to train survey methodologists so that they add value to their organizations, whether these are in government, industry, academia or elsewhere. Recently, the explosion of interest in alternative data sources (e.g., online transactions. social media, sensor data) that might be cheaper to acquire than survey data and whose creation places no additional burden on the public beyond effort already invested for other purposes (e.g., searching, posting, traveling) has led some of us to reconsider the kind of training that is most appropriate for students seeking careers conducting social research. The thinking is that students should be able to work with a range of data from the most carefully designed to data that are found and re-purposed. While many of the tools we currently present to students work across different data sources, other techniques will need to be added to the curriculum: machine learning, web scraping, automated text analysis, data visualization, i.e., Data Science courses. The goal would be to create a new inter-discipliary field — "Social Data Science" — but adding Data Science courses like these to the curriculum will require hiring new faculty and forming new partnerships (for example with Data Science programs and schools of Information Science). Whether the content of these courses is appropriate off-the-shelf or needs to be tailored to social research applications is not yet known. And how we can integrate instruction about data sources whose error properties are well understood (survey data) with data sources whose quality and generalizability are still unknown (e.g., social media data) represents a major challenge pedagogically and ethically.

Race in the U.S.: Police, Politics and Prejudice

Panel organizer: Marjorie Connelly, *NORC at the University of Chicago*

      As Barack Obama's presidency draws to a close, the public's view of relations between whites and blacks in the United States has deteriorated. The high profile deaths of several African Americans at the hands of police officers in recent years have exposed sharp differences in the views of blacks and whites on a number of issues, particularly law enforcement. How much lack of opportunity and prejudice is still experienced by African-Americans? How do perceived threats to whites' place in society affect their political views? These issues will be examined in papers by several distinguished pollsters: representatives from the Gallup Poll, the Pew Poll, the New York Times/CBS News poll, the AP-NORC Center poll and the University of Michigan. Darren Davis, director of the Center for Social Research at the University of Notre Dame will serve as discussant.

In the Twilight of the Obama Administration, How are Race Relations in the United States?
Megan Thee-Brenan, *The New York Times*

Most Americans say race relations in the U.S. are bad and not improving, and black Americans hold a particularly negative view of the nation's racial climate -- the worst since the country's first black president took office in 2009, according to the latest New York Times/CBS News poll. Following Barack Obama's historic win in 2008, both black and white Americans were notably sanguine about race relations in the nation. This marked a striking reversal among black Americans. In the run-up to the election, 6 in 10 blacks had characterized race relations as bad, however just several months into the Obama presidency, 6 in 10 blacks rated them as good. Since then, New York Times/CBS News polls have tracked deterioration in the public's perception of racial harmony. Still, for all of their negative feelings, black Americans do not appear to blame Mr. Obama for the breakdown as nearly three-quarters approve of his handling of race relations. In contrast, just 40 percent of whites approve of Mr. Obama's handling of the issue. While most whites and blacks agree that race relations in the US are generally bad, majorities of both races overwhelmingly say relations in their own communities are generally good. And whereas strong majorities of whites and blacks think most people are uncomfortable having conversations with people of another race, majorities of each race say they personally feel comfortable doing so. On the whole, many of the experiences and views of white and black Americans are vastly different, signs of a notable and persistent isolation of races in America.

The Divide between Black and White Americans: A Survey on Law Enforcement, Violence and Race
Dan Malato, *NORC at the University of Chicago*
David Sterrett, *NORC at the University of Chicago*
Marjorie Connelly, *NORC at the University of Chicago*
Trevor Tompson, *NORC at the University of Chicago*
Jennifer Benz, *NORC at the University of Chicago*

Many incidents of violence by police officers against civilians have made headlines in recent years and issues of police violence and the criminal justice system have become salient in the news. Critically, much of the police violence in the news has been against African-Americans, leading to discussions about the role of race in police work and potential biases in courts. These discussions raise several important research questions: In what ways do whites and blacks view police behavior – especially violent behavior – differently? How do they differ in what they attribute that violence to? Do they see it as racially motivated? What solutions do they see to reduce that violence? In 2015, the AP-NORC Center surveyed 1,233 adults, including an oversample of 311 African-Americans, to find out where the public stands on issues of police violence, the racial components of such violence, and potential solutions to the problem. This analysis shows that stark differences emerge in the rates that blacks and whites trust police, the degree to which they see police violence against civilians as a problem and one motivated by race, and whether they view the justice system as often acting too leniently toward violent officers. Using Census data, this analysis also shows how neighborhood diversity impacts views on these issues, with differences between blacks and whites becoming more moderate in more diverse neighborhoods, even if those racial differences are far from eliminated entirely. Still, the survey finds agreement across racial groups on many of the causes of police violence and broad consensus among the public on a number of policy changes that could reduce tensions between minorities and police and limit violence against civilians.

Interests, Identities and Hierarchy: The Central Role of Group Position in Explaining White Racial Attitudes

Vincent L. Hutchings, *University of Michigan*
Ashley A. Jardina, *University of Michigan*
Robert Mickey, *University of Michigan*
Hanes Walton, *University of Michigan*

What role do racial group loyalties and considerations of group interests play in structuring political preferences? The literature provides at least two answers to this question. Studies have shown that social groups exert significant influence on their members' political views. However, researchers have also found that White racial identity has only weak effects on public opinion. We reconcile these findings by extending the group position model of prejudice. We hypothesize that highlighting threats to the dominant status of Whites vis-à-vis salient out-groups are a necessary precursor to the activation of White identity. Results from our experimental design indicate that when Whites learn that their group may be losing ground relative to African Americans, but not Hispanics, White identity is more strongly associated with a variety of racial and political attitudes. Additional analyses on the 2012 ANES provide support for our contention that threat cues can activate White identity.

Public Views of Racism and Inequality
Jocelyn Kiley, *Pew Research Center*

Drawing upon several surveys conducted by the Pew Research Center, this paper will explore Americans' perceptions of racism and racial inequality. A 2015 survey found a growing sense from the public that racial inequities exist and should be addressed, and that racism is a problem in American society. These shifts are seen across racial, age, and partisan lines. Still, there remain significant divides in these attitudes: for instance, fully 86% of African Americans say the country needs to continue to make changes to give blacks equal rights with whites, while a narrower majority (53%) of whites say the same—and partisan differences are similarly pronounced. Using this 2015 survey, the paper will detail the trajectory of these views in recent years, and investigate their key correlates. The paper will also draw on additional Pew Research Center surveys, allowing for more in-depth analysis of views about race, racial equality, and perceptions of racism and discrimination within racial groups.

Americans' Re-assessment of Race Relations
Jeffrey M. Jones, *The Gallup Organization*
Lydia Saad, *The Gallup Organization*

Americans' assessments of black-white relations, including by blacks and whites themselves, turned sharply negative between 2013 and 2015. At the same time, Americans' views of relations between other combinations of the major U.S. racial and ethnic groups were unchanged. Also, the public's outlook for black-white relations going forward was also stable. Blacks' self-reports of personal discrimination were also largely unchanged. All this suggests the more negative assessment may be a temporary reaction to the prominent media attention to a series of fatal incidents between white police officers and black men in 2014 and 2015. The trends are part of Gallup's Minority Rights and Relations survey, which has been conducted 10 times since 2001 with black and Hispanic oversamples. Some trends on many items included in the survey date further back than 2001.

Using Response Propensity Modeling for the Cost-effective Allocation of Survey Incentives

Panel organizer: Paul J. Lavrakas, *Independent Consultant*

This panel will be comprised of four papers/presentations and a discussant. The topic of the panel will be the innovative use of Respondent Propensity Modeling (RPM) to determine the most cost-effective way to allocate noncontingent incentives in surveys in order to minimize differential nonresponse and thereby reduce the likelihood of nonresponse bias. By "most cost-effective," we mean the use of the finite resources available for incentives in a survey so as to maximize their positive effects on differential nonresponse and nonresponse bias. In this manifestation of RPM, a two-stage process is used whereby the most efficient statistical model is identified in the first stage. This model generates a propensity-to-cooperate score for each person/unit in survey's initially designated sample. This is done prior to the start of data collection. Once identified, the model is used to make decisions about how to best allocate the finite amount of funds that are available for incentives in the subsequent survey. In Stage 2, an experimental design is used to test the effects of the differential incentive treatments that were chosen based on the predictions of the model that was identified in Stage 1. The panel will include papers that address (1) the theoretical rational for the RPM approach as linked to Leverage-Saliency Theory, (2) different statistical techniques to create the model, (3) considerations that are used to determine the optimal differential allocation of incentives, (4) creating randomized experiments to test the differential RPM-allocated incentives, and (5) the results of two experiments using the RPM approach – one in a cross-sectional survey and the other in a panel survey.

Conceptual Background on Response Propensity Modeling for Allocating Differential Survey incentives: Purpose and Rationale
Paul J. Lavrakas, *Independent Consultant*

Traditionally, the same amount of a survey incentive has been given to all persons/units in a survey's initially designated sample. However, Lavrakas (e.g., 2014) has noted that this makes no theoretical sense, even if it does make the distributions of incentives a simpler task. As Groves, Singer, and Corning (2000) explained, Leverage-Saliency Theory posits that for every sampled respondent there is a different set of factors that come into play at the time s/he is made aware that s/he has been invited to participate in a survey. Some of these factors motivate the person to cooperate, whereas others motivate her/him not to cooperate. The balance between factors leads to the decision a given respondent makes about cooperating. For some respondents, incentives are a very important motivation to cooperate, whereas for other respondents incentives play little or no role in motivating them to cooperate. Following from this, a finite pool of incentive funding for a survey should be allocated across respondents so that those most likely to be positively affected by them get higher material incentives and those least likely to be affected get lower or no material incentives. This presentation will explain the theory and wisdom underlying the use of differential incentives to achieve the most cost-effective impact on differentially raising survey response with the goal of lowering nonresponse bias. A two-stage approach to identifying, applying, and testing a statistical model to predict each sampled respondent's propensity to cooperate in a subsequent survey will be discussed. This discussion will include (a) the types of data that can be assembled to use in the modeling, (b) an explanation of different statistical analyses that can be conducted to identify the best model, and (c) how the findings can be used to design experimental conditions to test the efficacy of the model.

Developing and Validating a Response Propensity Model for the Efficient Allocation of Non-contingent Incentives Using 2014 NHES Data
Michael Jackson, *American Institutes for Research*

The National Household Education Survey (NHES) is administered periodically by the National Center for Education Statistics using an address-based sample (ABS) and a mailed, self-administered questionnaire. The 2016 administration will incorporate an experimental test of a tiered incentive structure, in which different households receive a different value of a prepaid cash incentive in an amount determined by a predicted response propensity (RP) score. All this is determined prior to data collection. In preparation for this experiment, pre-collection research used the results of a 2014 NHES pilot study to develop, validate, and estimate a logistic regression model, which was then used to assign predicted RP scores to the 2016 sample. The 2014 data also were used to define the criteria by which a household's incentive amount was assigned in 2016. This paper reports the results of this pre-collection research and model building stage. It describes the auxiliary variables available for use in modeling, the analyses conducted to select a subset of these variables for inclusion in the final model, and the use of a cross-validation procedure to evaluate the final model's out-of-sample predictive power. It then discusses the use of this final model to identify an optimal incentive structure for the NHES 2016 experiment—namely, the RP score "cutoff points" that were used to determine a particular household's incentive amount. Finally, it compares the distribution of predicted RP scores in the 2014 sample to those in the 2016 sample to assess the stability of the model's predictions across administrations. The results provide evidence that a fairly parsimonious logistic regression model, estimated using auxiliary data appended to the ABS frame, generates in-sample and out-of-sample predictions that are sufficiently accurate and stable for use in assigning prepaid incentive amounts prior to data collection.

Can Response Propensity Modeling be Used to Improve Response Rates, Reduce Nonresponse Bias and Reduce Cost through the Use of Tailored Differential Incentives in the 2016 NHES
Cameron B. McPhee, *American Institutes for Research*

Incentives have long been used to improve response rates in surveys, and research suggests that in some circumstances incentives can reduce nonresponse bias by eliciting cooperation from reluctant or hard-to-reach individuals who are "different" on what is being measured from those easy-to-reach individuals. Incentives have also been shown to reduce total survey costs by reducing the level of effort required to gain cooperation. The 2016 administration of the National Household Education Survey, an address-based mail survey of 206,000 U.S. households, will include an experiment designed to test a tiered (differential) incentive structure, in which households will receive a prepaid cash incentive in an amount determined by a predicted response propensity (RP) score assigned prior to data collection. The tiered incentives are being assigned such that cases with the highest predicted propensity to respond would receive no ($0) noncontingent incentive, with the next most likely group receiving $2, followed by a group receiving $5, and finally the predicted cases predicted to be least likely to cooperate receiving $10. Two randomly assigned control groups are included where all cases, regardless of their RP, received either $2 or $5. This analysis will examine the impact of the incentive experiment on response rates and level of effort. These dependent variables for the modeled treatment groups will be compared to the control groups. Additionally, response rates within each RP group will be compared between those receiving a targeted incentive and those receiving a uniform incentive, in order to determine how the tailored incentive impacts response across different response propensity levels. Finally, appended auxiliary data will be used to examine the demographic distribution of respondents from both the

modeled and control treatments, to determine whether the modeled incentive protocol leads to a more representative respondent sample compared to a uniform incentive protocol.

Using a Response Propensity Model to Allocate Non-contingent Incentives in a Web Panel
Mark Masterton, *American Institutes for Research*

This presentation will report on efforts to use a response propensity model to target prepaid incentives in a repeated cross-sectional web survey of college students in Nebraska. The response propensity model will use self-reported respondent demographic data and institution-provided record data – both available on the frame and from prior waves of data collection – to predict response behavior in the administration of the new survey wave. We will then use the response propensity model results to design an experiment using differential incentives that will be implemented in the subsequent wave of this web survey. Sample members will either be assigned to the "standard (uniform) incentive condition" ($20 prepaid) or to one of the targeted (differential) incentive conditions. These differential incentive groups, which will be identified using the RP modeling approach, will be assigned to receive either $35, $20, or $10 prepaid incentive based on the results of the response propensity model. The presentation will first discuss the response propensity model, including the variables included in the model. The focus of the presentation will be the analysis of preliminary response rate data. This analysis will focus on response rates within each of the incentive groups, and the comparison of response rates and incentive cost per complete between those receiving a targeted incentive and those receiving a uniform incentive. We will also use demographic data available on the frame to compare the demographic distribution of respondents from both the experimental and control treatments.

If You Advertise It, Will They Come? Social Marketing, Digital Ads and Web Response

Panel organizer: Nancy Bates, *U.S. Census Bureau*

As traditional modes of telephone, mail, and in-person surveys continue to experience declining response rates and rising costs, internet surveys are gaining popularity as a cheaper and faster alternative. This fact is not lost on the U.S. Census Bureau. In 2013 the agency began offering the internet as a primary mode of response for the American Community Survey. For the 2020 Census, the agency has set a goal to enumerate 47.5 percent of households previously counted by mail instead via the internet. How best to offer and encourage internet response has been the subject of empirical studies. Thus far, the agency has relied mostly on a "push" method where households are mailed a paper invitation containing a URL and unique survey ID. In this session, we explore new methods to encourage web response. We report results from a site test where the use of preregistration, social marketing (including radio, TV, print ads and community partnerships) targeted digital ads, and different message content were manipulated to drive web response. We conducted a series of experiments to address the following research questions: § Can a pre-registration campaign elevate web response among "uber-respondents"? § How accurate is the delivery of targeted digital ads to intended audiences? § What is the level of web response via digital ad portals (i.e., no mailings, no ID required)? § How do different spend levels impact participation rates? Among traditionally hard to count populations? § How does targeted versus non-targeted ads impact participation rates? Among traditionally hard to count populations? § Does web survey participation/click rates vary by message

content, graphics, placement or ad platform? § How does digital advertising interact with traditional mail contact methods? § Can "influencer" calls drive web response?

Simulating a Census Environment to Test Online Self-response
Monica Vines, *U.S. Census Bureau*

The U.S. Census Bureau plans to leverage the internet and online self response as a primary response option for the 2020 Census. The 2015 Census Test in the Savannah Designated Market Area (DMA) provided an opportunity to test how an integrated communications campaign can support online self-response, particularly using the Census Bureau's capabilities to process responses without a pre-assigned ID number. This capability creates the opportunity for a paperless decennial census, for which residents access the Census Bureau's website and complete their forms online. For the 2015 Census Test in 17 Georgia and three South Carolina counties, Census performed blanket traditional advertising, partnership development, and public relations outreach. Additionally, Census segmented the 106 domestic ZIP codes in the area into five geographic panels, with all households in the market assigned to one of the panels. Each panel received different targeted online advertising strategies and per-household spending levels. Quantitative and observational findings about the effectiveness of different digital strategies will help identify promising approaches for promoting the 2020 decennial census. This presentation will provide an overview of the 2015 Census Test including results from the Optimizing Self Response and non-ID processing operations. We will discuss the selection of the test site and outline the panel design used to test digital advertising strategies. This will include a discussion of key digital advertising concepts such as targeting and advertising bleed.

Can We Understand What Works in Multi-channel Outreach Campaigns by Using Dedicated Web-addresses?
Samuel Hagedorn, *Penn Schoen Berland*
Kiera Maureen McCaffrey, *Reingold, Inc.*
John Otmany, *Reingold, Inc.*

Large-scale surveys that encourage participation through multi-faceted communications campaigns, such as the decennial Census, have typically experienced challenges attributing positive impacts to specific communication efforts (Datta et. Al, NORC 2012). As data collection efforts increasingly move online, the use of tracking codes via URL hyperlinks can help enhance attribution to identify what outreach modes were successful at encouraging response, especially among specific populations of interest. This paper explores results from a 2015 Census Test to identify the proximal communication channels that resulted in survey submissions. These channels include both online and traditional offline outreach methods, including digital advertising, partner promotion, postcards, QR codes, and opt-in text messages. By connecting user-provided response information with outreach channel-specific tracking information, the authors provide a more nuanced view of how survey participants engage with the data collection efforts and which response channels are directly effective at boosting participation. Using the URL-based attribution, the authors analyze the performance of digital advertisements to directly increase participation in data collection efforts. Results from three small-scale digital tests are discussed, including use of concrete directions in advertising call-to-actions, the appeal of localized language, and the impact of different legal requirement phrases. These tests emphasize the importance of placement and audience selection in digital advertising and suggest further opportunities for research on applications for survey collection. Implications for security standards in the context of the federal government and participant privacy are also discussed.

Encouraging Online Response among Hard-to-Survey Populations: Digital Advertising and Influencer Calls

Nancy Bates, *U.S. Census Bureau*
Matthew Virgile, *U.S. Census Bureau*

Williams et al. (2014) argue that social marketing was critical to the success of the 2010 Census. The 2010 campaign consisted mainly of traditional paid media including radio, television, and print advertisements, while only a fraction consisted of digital media. By 2016, however, online and mobile advertising are projected to become the world's second and third-largest ad mediums respectively, behind only television (Pomfret, 2015). In this paper, we report results of a site experiment that manipulated digital ad spending and targeting to encourage response among hard-to-survey populations. Additionally, among households where a phone number was located, a separate experiment tested three influencer call "voices" inviting online response (a Mayor vs local news anchor vs professional voice actor). We address a broad research question - namely, did digital advertising spend-level, ad targeting, or influencer calls increase online response among hard-to-survey populations? According to Tourangeau (2014), one can define hard-to-survey populations from a number of different perspectives including being hard to identify, reach, persuade, and/or interview. In this paper, we operationally define hard-to-survey based on tract-level population segments known to have low self-response in previous censuses (Bates and Mulry, 2011), stratifying areas by a Low Response Score (Erdman and Bates, 2014), and/or concentrating on priority audiences with historically low response and/or low internet usage (e.g. young adults, seniors, renters, and female headed households). In answering our research question, we will examine levels and mode of self-response among hard-to-survey populations (mail vs phone vs internet). For internet self-responders, we examine source (e.g. URL from mailings, digital ads, or non-digital ads), and finally, for those responding via digital ads, we examine type of device targeted (e.g. desktop versus mobile device).

Communication Channels that Predict and Mediate Self-response

Gina K. Walejko, *U.S. Census Bureau*

Past decennial censuses used an array of communication channels to build awareness and promote participation. The 2000 Census was the first to benefit from paid advertising. The 2000 media campaign included television, radio, Internet, print, and out-of-home advertising targeting both the general public and select populations. The 2010 Census campaign employed a bigger and expanded paid advertising operation. The 2020 Census will likely include an integrated communications campaign with more digital advertisements to complement an Internet self-response option and ability to process non-ID households in real-time. To examine the effects of advertising techniques made available by an Internet self-response option and non-ID processing, a mid-decade site experiment utilized a media campaign aimed at building awareness and promote self-response. Multiple communication channels included mail pieces, targeted digital advertisements, saturated advertising buys, telephone reminder or "influencer" calls, and emails or texts to pre-registrants. Using advertising data, paradata, and response data, this presentation examines which channels were most successful at predicting paper and Internet self-response modes. Using bivariate and multivariate analyses, we examine the effectiveness of different media campaign channels on self-response. We also measure the indirect, mediating effects of communication channels on self-response. Findings suggest different channels directly and indirectly affect self-response. While mail pieces directly affect self-response, there is evidence that other communication channels mediate these effects. We conclude by discussing how the results of this site experiment influence planning the 2020 Census' anticipated integrated communications campaign.

Tailored Recruitment for Specialized Samples

Novel Methodology for Reaching a Statewide Representative Sample of Youth Ages 12-18

Naomi Freedner-Maguire, *ICF International*
Randy ZuWallack, *ICF International*

The Washington Youth Marijuana Assessments project is a survey conducted with a representative sample of WA residents ages 12-18. Adolescents are a challenging population for survey researchers, and surveying a representative sample of adolescents brings additional challenges. As cell phones have become the primary mode of communication in many households, particularly those with children, interviewing adolescents by telephone has become even more difficult. As the concept of a household telephone continues to erode, we are faced with a more complicated and awkward process of obtaining consent and then interviewing the teen. For example, teens are frequently unavailable to take a survey at the time their parent has provided consent. But, parents are often not willing to provide r their child's cell phone number for researchers to contact them directly. Over 90% of adolescents use the internet daily, making web surveys a plausible alternative. However, there is no database of email addresses from which to draw a representative sample. To mitigate some of these challenges, we used an address-based sample (ABS) frame as an alternative to a dual-frame telephone survey. To reach our target population, we implemented a mail "push-to-web" strategy with an option to complete a mail survey. Because parental consent was required for an adolescent to participate, we placed parallel emphasis on recruiting both the parent and the adolescent. In this presentation, we review 1) our contact strategy, which included multiple mailings to recruit participants, use of pre- and promised incentives, and reminders in multiple modes; 2) the success of using a commercially available database to append emails to addresses; and 3) the comparative effectiveness of various options for parents to provide consent for their child: at the conclusion of the parent web survey, by responding to a reminder e-mail, or by calling a toll-free interactive voice response line.

Introducing Inbound Call Survey: A New Sampling Methodology

Burton Levine, *RTI International*

Inbound Call Survey (ICS) is a new sampling methodology. In ICS, individuals who place a telephone call which fails to connect to the intended number are recruited participate in a survey. There are three sources for the calls that fail to connect to the intended number: misdials, incomplete calls and disconnected calls. Telecommunications industry experts conservatively estimate that U.S. and Canadian telephone providers route in excess of 150 billion calls a month, 5 to 10 billion of which fail to connect to the intended number. Until now no one benefited from these calls. Our collaborator, Reconnect Research, has developed a technology that connects these calls to an interactive voice response (IVR) system that collects survey data. As a proof-of-concept, using ICS we obtained 6,926 competed interviews to a subset of the Behavioral Risk Factors Surveillance System (BRFSS) questionnaire. We present the results of a comparison of the demographic characteristics of the ICS respondents, BRFSS respondents and the population. The ICS respondents match the population demographic closely, much better than the BRFSS. And, we present the results of a comparison between ICS and BRFSS for several health outcomes.

Safety First: Ensuring the Anonymity and Privacy of Iranian Panellists' While Creating Iran's First Online Panel

Amir Farmanesh, *IranPoll.com*
Ebrahim Mohseni, *University of Maryland*

With thousands of active panelists from across Iran, IranPoll.com is the only online panel provider for the Iranian population, providing unprecedented and extensive CAWI capacity for researchers of the Iranian population. In line with ESOMAR/WAPOR guidelines, IranPoll.com takes its responsibility to protect the privacy and personal information of its panel members very seriously. Specifically, the guidelines call for "researchers must ensure that adequate security measures are employed to prevent unauthorised access, manipulation and disclosure to the personal data, including any possible third parties." To achieve this and ensure safety of our panelists, IranPoll.com has adopted a series of measures to give confidence to its Iranian panel members that their personal identity will never be disclosed to any third party. The methodological paper will discuss in detail the methodological and technical we have undertaken to ensure this high standard. These measure extend from standard privacy protocols, to providing respondents with high-quality VPN services that to a degree prevents online fingerprinting by masking their exact whereabouts, browser and system details, and IP addresses when they partake in our survey.

Building a City-wide Survey Panel: Engaging Philadelphians through Multiple Modes

Nina DePena Hoe, *Temple University*
Heidi E. Grunwald, *Temple University*
Keisha Miles, *Temple University*

Background: The Institute for Survey Research (ISR) at Temple University in Philadelphia, PA is creating the country's first attempt at a city-wide survey panel, representing the diverse demographics of Philadelphians. The goal of the panel is three-fold: (1) To create a cost-effective and convenient resource for investigators conducting social science or public opinion research in Philadelphia; (2) To create a civic-engagement tool where Philadelphians can have their voices heard through opting in to participate in polls and surveys pertaining to life in the City of Philadelphia, (3) To conduct survey experiments and better understand the nature of opt-in panel respondent behavior at the local level. Methods: ISR is recruiting residents using five methods: (1) address-based, door-to-door outreach, (2) a designated street team conducting community outreach, (3) advertisements on public transportation, (4) social media, (5) mailed postcards, and (6) promotion through existing ISR field studies in Philadelphia. Participants can opt in to join the panel using SMS text messages, through a web survey, or in person with a field staffer. Once deemed eligible, members provide demographic information and select their preference of survey-taking mode (SMS Text, Web, Phone, Paper) and frequency of participation. Results: Preliminary findings indicate that typically harder-to-reach populations are willing to opt in to a local survey panel. Thus far, 65% of panel members are African American and 38% have a high school diploma or less. Of significance, 40% of members prefer to take surveys via SMS text message and 90% want to take more than one survey per month. Contributions: Findings from this local panel creation process provide insight into the nuances associated with recruiting for a city-wide panel as a civic engagement tool, and highlights the importance of offering multiple modes of participation (particularly SMS text) to better access harder-to-reach populations.

Using Nonresponse Follow-up (NRFU) Recruitment to Help Build a Probability-based Research Panel

Nadarajasundaram Ganesh, *NORC at the University of Chicago*
J. Michael Dennis, *NORC at the University of Chicago*
Paul J. Lavrakas, *NORC at the University of Chicago*

Most internet research panels throughout the world are formed using nonprobability sampling methods. However, past research (e.g., Yeager et al. 2011, Chang & Krosnick 2009, Walker, Pettit & Rubinson, 2009) has consistently shown that probability-based panels provide more accurate results than those built using nonprobability designs. Funded and operated by NORC at the University of Chicago, AmeriSpeak is a probability-based panel in which the vast majority of panelists provide data via the internet, with a minority providing it via telephone interviewing. AmeriSpeak uses an area probability sample frame (based on address-based sampling) and uses multiple modes of recruitment, including mail, telephone and in-person. As a probability-based panel, AmeriSpeak is designed to be a nationally representative panel of the English-speaking and Spanish-speaking residential population of the U.S. The recruitment process for AmeriSpeak occurs in two phases: (1) initial recruitment; and (2) non-response follow-up (NRFU) recruitment. In the latter, a random subsample of initial nonrespondents is selected for in-person recruitment. This presentation discusses the sample and data quality impact of the NRFU recruitment for the AmeriSpeak panel. We will present data on the demographic and geographic representation of the AmeriSpeak panel by initial and NRFU panel recruits. We also will present findings showing the impact of the NRFU subsample on panel representativeness, design effect, and key attitude and opinion survey estimates. In addition, we will report the impact of NRFU recruitment on bias reduction and higher cumulative AAPOR response rates for panel surveys.

Evaluating, Monitoring and Improving Interviewers' Performance

The Use of CARI and Feedback to Improve Field Interviewer Performance

Aaron Maitland, *Westat*
Ryan Hubbard, *Westat*
Brad Edwards, *Westat*

Presenting respondents with a consistent stimulus is important to standardized survey interviewing. Relatively few studies have had appropriate research designs to investigate how interviewer behavior is related to data quality. There are a number of aspects of interviewer behavior that might influence the quality of data. For example, it has been shown that poor interviewer behavior leads to interviewer effects (Mangione, Fowler, and Louis, 1992) and decreased accuracy (Belli and Lepkowski, 1996; Dykema and Schaeffer, 2005). This paper presents an experiment that tests the hypothesis that using computer assisted recorded interviewing (CARI) to provide timely feedback to field interviews improves interviewer performance. Field interviewers conducting CAPI household screeners were randomly assigned to either receive relatively quick feedback on their first few interviews or to only receive feedback as part of standard validation procedures. In the treatment group, CARI recordings were behavior coded to assess whether questions were read exactly as worded and whether responses were recorded accurately. Interviewers received feedback, identifying specific problems and reminding them of the importance of reading questions exactly as worded. Interviewers were also provided with a direct link to the CARI recording and were encouraged to listen to the errors they made. Preliminary findings demonstrate that providing interviewers with quick feedback at the beginning of the field period improves interviewer performance relative to interviewers in the control group who

received less timely feedback. This finding is consistent across different types of interviews in the same study and across interviewers with different levels of experience. We also examine the implications of this finding for data quality.

Mitigating Impact of Interviewer Burden
Dakisha Locklear, *RTI International*
Tamara L. Terry, *RTI International*
Sabrina Bethea, *RTI International*
Kelly Castleberry, *RTI International*
Randy Spain, *RTI International*

Interviewer performance is crucial to successful data collection efforts. In spite of numerous technology-based advances with systems, tools, methods and metrics in recent years that have targeted and contributed to performance improvements, little focus has been placed on the issue of Interviewer burden.

The repetitive, detailed nature of interviewing, whether by phone or in-person, is known to be associated with fatigue, burnout and decreases in morale and engagement. Neuroscience research suggests that cognitive awareness, measurement and stimulation can help mitigate and/or control for such historically unavoidable factors. Multiple web-based platforms now exist for commercial and private usage that measure, track and improve individual cognitive abilities and options exist for customizing programs toward specific cognitive workloads. When coupled with Research suggests providing interviewers with anti-tedium objects can mitigate the negative effects of repetitive call center tasks, these new platforms may offer a method for improving interviewer performance and combating work-related fatigue.

The purpose of this experiment is to test the effectiveness of anti-tedium objects for mitigating burnout and fatigue in a call center environment. This data will be used to compare the effects of the cognitive engagement platform to a control group on interviewer performance and employee engagement throughout their shifts. By providing interviewers with the opportunity to use mind-stimulating activities throughout their shifts, we hope to demonstrate that interviewing fatigue can be reduced and lead to improved productivity and data quality.

Inside the Mind of the Interviewer: Applying Cognitive and Social Psychology to Interviewers
Erica C. Yu, *U.S. Bureau of Labor Statistics*
Robin L. Kaplan, *U.S. Bureau of Labor Statistics*

Over time, our understanding of the survey interview has evolved into a realization that respondents are regular people with flawed memories and unique interpretations of questions. However, the field of survey methodology has largely minimized the role of the interviewer to characteristics such as race, gender, personality, and attitudes. Just as we now understand that the task of answering questions spans comprehension, retrieval, judgment, and response processes, we need to acknowledge the cognitive complexity of the interviewer's task in the data collection process. This presentation will review our existing knowledge about interviewers and present a framework for understanding interviewer cognitive processes and their impact on data quality. We begin from the premise that interviewers are equally as vulnerable to cognitive and social biases as we now know respondents to be. We develop a model of the interviewer's task and identify the judgments and decisions that interviewers may make during data collection. We apply theories from a broad range of

cognitive and social psychology fields including other domains that use the interviewer paradigm, such as police interrogation and medical diagnosis, to hypothesize what factors lead to commonly observed interviewer behaviors. We use probing as an example: How do interviewers decide whether a response is adequate? Which probe will elicit honesty from the respondent? Is it worth it to probe when that might slow down the interview or frustrate the respondent? Factors that may influence these judgments and decisions include the interviewer's own understanding of the survey goals, beliefs about the respondent's true value for a given survey question, feelings of sensitivity induced by asking invasive questions, and the nature of any incentives, among others. An understanding of these processes should inform how we design surveys and train interviewers, in order to encourage ideal behaviors.

Monitoring Field Interviewers: Real-time Applications of Paradata in Experimental Designs

Rachael Walsh, *U.S. Census Bureau*
Scott Boggess, *U.S. Census Bureau*

Both household and establishment survey response rates have been declining over time. Additional effort is necessary to maintain response rates when using interviewers to collect data in-person. Through research and experimental designs in the field, the U.S. Census Bureau makes a concerted effort to reduce the cost of data collection while maintaining both response rates and data quality. After analyzing historic paradata from the National Ambulatory Medical Care Survey (NAMCS), a series of tests were designed to reduce the cost of data collection. One test restricted the number of personal visit attempts interviewers made to office-based physicians and community health centers (CHCs). The experimental design utilized past response rates and caseload distribution when determining the test and control group areas. Interviewers in the test group were restricted to six personal visits during the first two phases of the three-phase interview, and ten personal visits for completion of the full interview. During data collection, headquarters staff generated paradata metrics used to monitor interviewers by combining data from the sampling frame, the case management system, and the contact history of the case. To prevent noncompliance across interviewers in the test group, headquarters staff used these metrics to compile a list of cases approaching noncompliance, assisting field supervisors in the near real-time monitoring of interviewers during the data collection process. We evaluated the test for reductions to the cost of data collection while maintaining data quality. This application of auxiliary data during the data collection process monitored interviewers in near real-time, successfully piloting the personal visit restriction test where all interviewers in the test group complied with the test parameters. Though the cost of data collection did not decrease, the test group areas did have higher response rates than the control group areas, making the test a relative success likely for nationwide implementation.

Improving Interviewer Performance through the use of Modeled Paradata Reports

Tamara L. Terry, *RTI International*
Rebecca J. Powell, *RTI International*
Timothy Nesius, *RTI International*
Marion Schultz, *RTI International*
Howard Speizer, *RTI International*
Kelly Castleberry, *RTI International*

The caliber and performance levels of interviewers are pivotal to successful data collection. The more efficiently, consistently and comprehensively data collection managers can measure and assess

Interviewer-level performance-based factors, the more targeted and strategic managers can be with prioritizing training and the mentoring of staff. The Research Operations Center at RTI International has developed and implemented reports that use the case history and other paradata to account for sample differences across interviewers to allow for faster and more accurate detection of underperforming interviewers, using a metric similar to the one in West and Groves (2013). The results from the first experimental evaluation conducted on a large telephone survey proved a) reports using modeled paradata allowed for quicker and more accurate detection of interviewers who were not performing on a week-to-week basis, and b) although the results were not significant, the reports using modeled paradata had the most positive impact on overall interviewer performance (Terry, Peytchev, et al, 2015). In this paper, we will focus around how the modeled paradata reports are used weekly and long term to identify underperforming staff, the value of having a report that uses paradata to assess interviewer skills, and the enhancements we have made to the report that continuously allow us to improve interviewer performance. Our goal is to create an automated, standardized and robust reporting platform designed and built to essentially pull and compile pre-determined measurement factors and paradata from across various data collection efforts. We also plan to provide a brief comparison of the use of our modeled paradata reports to our standard interviewing performance feedback techniques. This paper hopes to provide much needed experimental support for devoting more effort to process and transform paradata, in order to make these data more useful during data collection.

Examining Election Polling Issues Cross-Nationally

Reshaping the Data Collection with Landlines: The Case of the Most Recent Elections in Mexico

Diana Paola Penagos Vasquez, *Parametria S.A. de C.V.*
Francisco Abundis Luna, *Parametria S.A. de C.V.*
Jose Alberto Vera Mendoza, *Parametria S.A. de C.V.*

Modern democracies can not be defined without full realization and dissemination of opinion polls. Changes in the coverage of public services, improvements in health, higher educational levels, higher levels of income and massive exploitation of new communication networks, have been shaping a more critical and better informed citizenry. Forums as WAPOR, MAPOR or AAPOR have put on the table the discussion about new ways of gathering information in accordance with these social transformations. Telephone surveys, the use of cell phones, e-mails and the internet are some of the modern mechanisms to collect population data. However, the development of these tools is uneven across countries. In developing countries such as Mexico, the telephone coverage is statistically associated to the level of household income. National Institute of Statistics and Geography (INEGI) notes that only 44% of the Mexican households have a landline. This is unavoidable methodological challenges due to widespread bias that such measure could represent. This research analyzes one of the most recent elections in Mexico where both sample design applied with totally contrasting data. By comparing information from both exercises with the final results of the election, we can say that there are specific cases where telephone surveys captures more adequately the preferences of voters. Undoubtedly, the methodological changes of this kind require in-depth analysis in countries where coverage bias seems to play against.

Mode Effects in Electoral Polls: A Comparative Perspective
Claire Durand, *Université de Montréal*

During the last decade, Web and IVR (Interactive Voice Response) polls have spread in many countries and particularly in small markets where media outlets do not have the resources necessary to sponsor electoral polls conducted using "classical" modes. In addition, there is no incentive to put money in classical telephone polls if IVR or Web polls perform well and cost much less. But is it really the case? In this presentation, I analyze the performance of Web and IVR polls in various contexts, namely recent Canadian elections at the federal and provincial levels, the Scottish referendum of 2014 and the US election of 2012. I examine poll performance according to mode of administration on two aspects: a) systematic error, i.e., the capacity to forecast election results without bias, and b) random error, i.e., the level of variability in estimates, taking into account the sample size. I show that Web and IVR polls tend to have systematic bias that may – and should – be corrected in order not to mislead voters. IVR polls also tend to show substantial random error in specific contexts. In addition, bias is not always present so that we need to be able to identify which context and which specific methodological features are more likely to lead to bias. I examine these questions and conclude on how to correct bias, if possible, and, on the necessity to inform voters about the possible biases of the polls they are consulting.

We've Got You Surrounded: Multiple Data Streams and the 2015 Canadian National Election
Darrell J. Bricker, *Ipsos Public Affairs*
Sean Simpson, *Ipsos Public Affairs*

Data collection for election polls in Canada use three main interviewing methods - computer-assisted telephone with live interviewers (CATI with dual frame sample), interactive voice recording (IVR), or online (both panel and river-based). Ipsos collected Canadian vote intention data using two of the three methodologies, specifically dual-frame CATI, and online (panel, river). This allowed us to isolate the effects of each data collection method while controlling for script and schedule impacts. Our results show significant differences in the measurement of vote intention between the methodologies. Online methodologies over-represented NDP voters, while telephone methodologies over-represented CPC voters. The LPC vote was measured consistently by both methodologies. After reviewing the test results, the authors then suggest some weighting adjustments that pollsters should consider to offset the potential measurement biases inherent in both online and telephone data collection.

Observations on Improved International Election Polling: Learnings from Opinion Research During the Greek Bailout Referendum Election
Chris Jackson, *Ipsos Public Affairs*
Clifford Young, *Ipsos Public Affairs*

The July 2015 Greek Bailout Referendum Election presented unique challenges for opinion research. From Tsipras's announcement to Election Day spanned only 8 days leading to a very compressed schedule and an unsettled electorate where "conventional" approaches to election polling fell short. This paper will present original data and analysis from the Greek Referendum Election to show a superior method of collecting accurate vote projections in low-information environments. Ipsos Public Affairs, with Eurasia Group and Ventris Research, conducted telephone tracking polling over the 6 days preceding the election. Over the course of the poll, Ipsos used four calibration methods to represent the electorate. The first approach was to weight data to census demographics – standard for opinion

research. The second calibration was weighting data to the previous election result – the standard polling approach in Greece. The third approach was using a fixed likely voter formula to identify voters – normal practice for U.S. elections. The forth calibration was using a dynamic likely voter screen adjusted for expected turnout – Ipsos Public Affairs approach to U.S. election. This polling data and the actual election results illustrate that the Ipsos Public Affairs approach most accurately predicted the final election results.

Positive vs. Negative: How Question Polarity in Voting Advice Applications Affects the Answers and the Voting Advice
Bregje C. Holleman, *Utrecht University*
Naomi Kamoen, *Utrecht University*
Jasper van de Pol, *ASCOR/University of Amsterdam*
André Krouwel, *Free University*
Claes C. de Vreese, *ASCOR/University of Amsterdam*

Online Voting Advice Applications (VAAs) are survey instruments that help citizens to shape their political preferences and compare them with those of political parties. Especially in multi-party democracies their increasing popularity suggests that VAAs play an important role in opinion formation for individual citizens, as well as in the public debate prior to elections. Hence, objectivity and transparency of VAAs are crucial. In the design of VAAs, many choices have to be made. This study investigates whether question polarity affects the answers to VAA statements and the voting advice: do positive statements (e.g., the city council should allow cars into the city centre) lead to different answers and to different voting advice from negative ones (e.g., the city council should ban cars from the city centre)? Extant research on the so called forbid/allow asymmetry suggests an overall effect of positive vs. negative wordings. This study is the first to investigate polarity effects on a much wider range of contrast types within one study. In a field experiment (N=31,112) during the Dutch municipal elections we varied the polarity of 18 out of 30 VAA statements. Analyses show a significant effect of question wording for two-thirds of the manipulated questions, but not one overall effect for positive vs. negative questions. This large variation in wording effects between different questions and contrasts is explored further in several post-hoc analyses. Despite the variation in the size and direction of wording effects, a clear relation with the voting advice is shown. These effects of question polarity are not confined to users with lower levels of political interest, but rather occur for a broad range of citizens. This stresses the need for VAA builders to be sensitive to wording choices when designing VAAs, especially since their voting advice is known to affect citizens' vote choice.

Measuring Lifestyles: Sleeping, Drinking, Gambling and Parenting

The New Living Room: How Americans View their Video
Julie Phelan, *Langer Research Associates*
Gary Langer, *Langer Research Associates*
Wendy Loxley, *ESPN*
Flora Kelly, *ESPN*

In partnership with ESPN, building off groundbreaking qualitative research and using the probability-based GfK KnowledgePanel, we conducted an extensive survey assessing the viewing patterns and preferences of 13- to 54-year-old Americans who had streamed video in the previous week. Our analysis includes "preference x priority" scores assessing the relative strengths and

weaknesses of streaming vs. cable; an evaluation of how, why and when viewers switch between the two modes; regression analyses determining the strongest independent predictors of each viewing activity; and a cluster analysis identifying key video viewing groups. Results show crucial attitudinal, behavioral, generational and other demographic differences between watching cable or broadcast TV vs. streaming. The 2016 AAPOR conference is devoted in part to examining current changes in substantive public opinion. Our paper evaluates a dramatic change in a prominent public behavior, analyzing its motivators as content providers seek to adapt successfully to the streaming-fueled future. With its rigorous design and analysis, the study also demonstrates the practical value of high-quality survey research techniques in an applied setting.

What Do You Mean I'm Drunk?! Development and Testing of a Conceptual Framework for Asking About Intoxication

Nina Van Dyke, *Social Research Group (a division of Market Solutions)*
Anna L. Lethborg, *The Social Research Centre*
Julaine Allan, *Social Research Group (a division of Market Solutions)*
Christine M. Maddern, *Market Solutions*
Sean O'Rourke, *Victorian Responsible Gambling Foundation*
Emma L. Saleeba, *VicHealth*

VicHealth, a public health foundation in Victoria, Australia, is engaged in enacting an alcohol cultural change program that encourages a healthy drinking culture among young people aged 16 to 29. As part of this initiative they asked us to identify and develop indicators of alcohol culture with a focus on acceptability of intoxication. A literature review showed that existing survey questions either: (a) ask about alcohol consumption as a proxy for intoxication, or (b) provide terms such as "drunk" or "intoxicated", assuming that respondents agree on what these terms mean. Using a mixed methods approach, we first conducted qualitative research to understand how young people talk about the alcohol culture and intoxication. Key findings were: (1) young people discuss intoxication in terms of behaviours, not consumption; (2) young people use a given term for intoxication to mean quite different levels of intoxication behaviours; (3) young people generally agree on stages of intoxication as described by behaviours. These findings were used to develop a new approach to asking survey questions about intoxication, which was trialed in a survey of almost 1500 Victorians. This new approach consists of the following steps: (1) ask respondents to rate intoxication behaviours on a 0 to 10 scale; (2) ask respondents to choose their own term for "losing your balance", a behaviour consistently rated towards the higher end of the scale (mean of 8.1); (3) ask respondents questions about acceptability of intoxication using their chosen term. This approach thus anchors personal perceptions of intoxication to an objectively identifiable behavior consistently associated with a particular (high) level of intoxication. We believe that this approach is superior to previous methods and has the potential to be applied to other types of behaviors, although further testing is needed.

The Sleep Health Index: How America Gets its Zzzzs

Julie Phelan, *Langer Research Associates*
Gary Langer, *Langer Research Associates*
David Cloud, *National Sleep Foundation*
Kaitlyn Whiton, *National Sleep Foundation*
Kristen Knutson, *University of Chicago*
Gregory Holyk, *Langer Research Associates*

Sleep is a subject of deep fascination but limited measurement – a troublesome shortfall, since the Centers for Disease Control identifies insufficient sleep as a public health problem affecting 50 to 70 million Americans. We report on a project produced in conjunction with the National Sleep Foundation that seeks to quantify and track Americans' sleep habits on an ongoing basis – the national Sleep Health Index. We used theoretical and empirical methods to produce the index. Questions administered in nationally representative telephone surveys in 2014 and 2015 were selected for inclusion on theoretical principles and tested for reliability and validity using statistical analyses. Factor analysis established that the 14 index questions relate to three discrete constructs – sleep duration, sleep quality and disordered sleeping. Computed on a 0-100 scale, these were assembled as subindices and combined to form the overall SHI. We tested the subindices and full index for internal consistency, reliability over time, convergent validity and known-groups validity, with results showing empirical validity and consistency. We also produced a regression analysis to find the strongest independent predictors of sleep health. Stress levels and overall health are most important, with positive predictors including life satisfaction and a comfortable mattress, while negative predictors include snoring, being older and using electronic devices after bedtime. 2015 scores were 74 overall (the equivalent of a "C" letter grade), with 82 for non-disordered sleep, 72 for sleep duration and 69 for sleep quality. (Higher scores reflect greater sleep health.) Index grades were well-distributed; 11 percent of Americans received an A grade for their sleep health, 36 percent a B, 22 percent a C, 13 percent a D and 17 percent an F. We'll also discuss future work, including use of the index as a predictor of other outcomes.

Hitting the Jackpot on Survey Response: The Relationship Between Gambling Behavior and Current and Proposed Casino Proximity

Katie Dekker Archambeau, *NORC at the University of Chicago*
Ned English, *NORC at the University of Chicago*

The Expanded Gaming Act was signed into law in 2011, allowing Massachusetts to introduce casinos and slot parlors for the first time. As a condition of this legislation, the Massachusetts Gaming Commission was required to conduct research to understand the effects of expanded gaming and to minimize any harmful effects of gambling. The University of Massachusetts-Amherst and NORC at the University of Chicago partnered to conduct a survey of the general population in Massachusetts prior to the introduction of gambling. The Social and Economic Impacts of Gambling in Massachusetts (SEIGMA) survey inquired about several topics including health status, gambling attitudes and motivation, prior year gambling behavior, and gambling problems. A subsample of SEIGMA respondents were selected to participate in the ongoing Massachusetts Gambling Impact Cohort (MAGIC) Study in an attempt to track gambling behavior over time. This study represents a unique opportunity to survey residents prior to the introduction of gambling, allowing researchers to assess long term impacts. SEIGMA was implemented as a statewide general population survey. This presentation will analyze responses to the SEIGMA survey as related to the geography of the survey respondent. Our research examines survey response rates, problem gambling rates, and awareness of media campaigns by geographic proximity of a respondent to

current out-of-state gambling facilities and future sites of licensed Massachusetts facilities, in addition to extant demographic information. Additionally we compare these same metrics between SEIGMA and the first wave of MAGIC to look for changes in response by geography. The results of this analysis will provide insight into a person's propensity to respond to a survey as it relates to relevancy of the survey topic, in this case measured by distance between a participant and a gambling facility.

Methods to Improve Data Collection Efficiency

Designing Efficient Field Locating Strategies: Local Staff, Travelers or Both?
Brian H. Roff, *Mathematica Policy Research*
Stacie Feldman, *Mathematica Policy Research*

Face-to-face data collection, part of which may include field locating, remains a frequent means of obtaining primary data despite significant costs associated with recruiting, training, and managing field staff. To reduce costs, survey organizations often recruit individuals who live in or near sample areas instead of, or in combination with, individuals who have to travel. Yet, because "travelers" often have many years of experience and are deemed reliable, they often support or replace local staff near the end of the field effort when and where response rates are low. A critical question is whether this combined use of local and traveling field staff is the most productive and cost effective way to conduct field locating? We examined this question by randomly assigning 7 local and 5 traveling field locators to clusters of nonrespondents in four metropolitan areas as part of a CATI survey with field follow-up. The survey involved interviewing participants 18 months after enrollment in a jobs training program with a fielding phase implemented over multiple 8 week periods. The experiment was conducted during the first 8 week period. Travelers were randomly assigned to metropolitan areas; then, both traveling and local field locators were randomly assigned to a cluster of cases of equal size and similar geographic distance between cases. Field locators were instructed to attempt contact with all assigned cases within 12 days. We found that over the entire 8 week period, local staff performed as well as traveling staff in gaining response, though differences in contact and completion rates were notable during the first 12 days. This paper discusses: (1) differences in performance outcomes by field staff type and over time; and (2) implications for structuring future survey field efforts.

Pre-testing Establishment Surveys: Moving Beyond the Lab
Heather Ridolfo, *National Agricultural Statistics Service*
Kathy Ott, *National Agricultural Statistics Service*
Jeremy Beach, *National Agricultural Statistics Service*
Jaki S. McCarthy, *National Agricultural Statistics Service*

Cognitive interviewing is a widely used method for ensuring that survey questions produce valid data. However, it is also a very time consuming and expensive method. In household surveys, cognitive interviewing is typically conducted by experienced survey methodologists in lab settings. Agencies with labs can conduct multiple cognitive interviewers in a day using small staffs of experienced methodologists, thus reducing costs and expediting the research process. But when populations are hard to reach, methodologists must travel to participants. In surveys of establishments, cognitive interviewing is almost always conducted outside of a lab setting. Researchers must travel to multiple establishment sites that are not always geographically close to one another. At the National Agricultural Statistics Service (NASS), we have been able to do limited cognitive interviewing due to the rural nature and geographic distribution of its survey universe, namely, farm operators. Conducting interviews in

rural areas presents expensive and time consuming logistical concerns, including recruiting and setting up interviews in a short amount of time, and the expenses in both money and time of sending headquarters staff across the country. NASS expanded its cognitive interviewing program last year by training non-survey methodology staff members in state and regional offices to conduct cognitive interviews, under the direction of the survey methodologists at the headquarters offices. We will present information on how we conducted our expanded cognitive interview training program, our experiences so far, lessons learned, and how we plan to continue using our non-traditional cognitive interviewing program.

Using Text Analytic Techniques to Create Efficiencies in Analyzing Qualitative Data: A Comparison between Traditional Content Analysis and a Topic Modeling Approach

Andrew Stavisky, *U.S. Government Accountability Office (GAO)*
Philip Resnik, *University of Maryland*
Chris Musialek, *University of Maryland*

The U.S. Government Accountability Office (GAO) conducts independent, non-partisan policy studies for Congress. Many of the studies include qualitative research components, such as case studies, in-depth interviews, focus groups, and expert forums. The methodology for analyzing qualitative data for these studies typically uses comprehensive content analysis, an extremely time and resource-heavy approach. Therefore, GAO has begun to explore ways to find efficiencies in the analytic process while still maintaining the rigor and quality of the analyses. Topic modeling refers to a family of computational techniques that extract latent topics or themes from collections of text, bringing to the surface underlying structure that is not likely to be apparent via traditional techniques such as term-frequency analysis. Variations of topic models have been used in the social sciences for diverse purposes, ranging from analyzing trends in the scientific literature, to characterizing patterns of agenda-setting and agenda-framing in political communication, to identifying thematic categories in psychotherapy transcripts. In this study, traditional content analysis was compared to a topic modeling approach. The data for this study comprised of 12 focus group transcripts from a completed GAO investigation. The GAO investigation used a comprehensive content analysis to analyze the data. Recommendations to a Federal Agency about both cost savings and safety improvements were made based on the content analysis from these focus groups. Using the same original data set from the GAO investigation, we used a topic modeling methodology, combined with an interactive process involving human domain expertise, to determine whether efficiencies in the analytic process could be gained and to assess the effects of our combined human-machine process on quality and rigor. Our findings show promise for the use of a topic modeling approach, which has implications for both public and private-sector firms that use qualitative methods such as focus groups and in-depth interviews.

Optimal Routing for Field Interviewers: A GIS-based Approach for Reducing Household Survey Costs

Brad Edwards, *Westat*

Data collection in the field is almost always the cost driver for in-person household surveys. The largest share of the interviewer's time is spent traveling from home to the assigned sampled units, and from unit to unit. The field interviewer routing issue is similar to the traveling salesman problem, but is also akin to the scheduling function in a telephone call center. (In a call center, sample cases are served up to interviewers in an order based on a number of factors, such as time zone, propensity to respond at

specific times, staff availability, appointments, callbacks, new work, etc.) The means to address this two pronged problem have been lacking until recently. Today, analytics based on GIS data and paradata such as call attempt outcomes from individual cases enable managers to see much more clearly what is happening in the field in near real time, and to take corrective action. We report on experiences from two national surveys that have provided interviewers with mobile devices equipped with GPS capability. Interviewer routing and visits to specific units are observed by supervisors. Paradata analytics inform judgments about optimal case work and routing. Supervisors review the choices interviewers made about cases to visit and routes taken. The supervisors discuss the actual routes with the interviewers that afternoon or the next day, and make suggestions for more efficient routing, including use of navigation functions. Routing behavior during the next work block is observed, and additional feedback is given as required. We discuss the impact of this new capability on field survey efficiency and costs.

May 13, 2016
Concurrent Sessions D

Professional Development Panel: Project Management: Lessons Learned from Experienced Project Directors

Panel organzier: Rachel A. Caspar, *RTI International*
Tom Smith, *RTI International*
Jennifer Hunter Childs, *U.S. Census Bureau*
Bill Davis, *Davis Research*
John Stevenson, *University of Wisconsin Survey Center*
Eran N. Ben-Proath, *SSRS*

This panel will bring together experienced project directors to share tips for successful survey project management. The panelists will discuss a variety of topics including: staffing the project team; roles and responsibilities; effective communication; balancing the competing demands of budget, schedule and quality; monitoring the budget; error detection and remediation; managing "scope creep"; priority-setting; and meeting (or exceeding!) client expectations. Panelists will share examples from their work that highlight approaches that worked well and resulted in successfully completed projects. Panelists may also share experiences they've had with projects that did not run smoothly and, with the benefit of hindsight, what they might have done differently. Their goal in sharing these examples will be to help others to avoid making the same mistakes.

Alternative Approaches to Increasing Response Rates

Increasing Cooperation in Telephone Surveys with the Progressive Engagement Technique: An Australian Perspective

Paul Myers, *The Social Research Centre*
Paul J. Lavrakas, *The Social Research Centre*
Darren Pennay, *The Social Research Centre*
Natasha Vickers, *The Social Research Centre*

The introduction to a telephone survey has long been recognised as a critical factor in gaining cooperation. The conventional approach to introducing surveys to respondents sees interviewers being asked to deliver a standardised statement about the purpose and sponsor of the survey before inviting a reply; typically through asking the sample member if they would be willing to take part. By contrast, the progressive engagement technique seeks to create a two-way interaction between the respondent and interviewer through explicitly requiring and inviting the respondent to answer and ask questions immediately after making contact (cf. Burks, Camayd-Freixas, Lavrakas, & Bennett, 2007). Through creating a more natural and respectful style of interaction, the progressive engagement technique is thought to significantly increase willingness to take part in telephone surveys. We will present findings from an Australian experiment of the progressive engagement technique. The experiment seeks to replicate recent research by the NORC at the University of Chicago in the context of the Australian National University (ANU) Poll – a quarterly public opinion survey of Australians conducted via a dual frame RDD design by the Social Research Centre, using approximately 9,000 phone numbers. The experiment will involve randomly assigning members of the interviewing team to one of two conditions: implementing the progression engagement technique or implementing the conventional survey introduction. Further, the RDD telephone numbers that are sampled for the survey will be randomly assigned to one introduction condition only. Thus, all numbers assigned to the progressive engagement condition will only be processed with the progressive engagement introduction and only by interviewers trained in that approach; the same will be the case for the numbers assigned to the conventional introduction. Key outcome measures that will be examined include response and cooperation rates, field efficiency, interview length, item-level non-response and respondent profile.

Does Embedding a Survey Question in the Survey Invitation E-mail Affect Response Rates? Evidence from a Randomized Experiment

David L. Vannette, *Qualtrics*

Web survey invitations are most commonly sent to potential respondents via an e-mail request for participation. The flexibility and rapid innovation possible in e-mail and online survey technologies means that novel approaches for recruiting respondents are continually being developed. In much of this research, the most important outcome metric evaluated is the survey response rate. Low survey response rates are an important driver of increased costs and longer field periods in survey data collection. Identifying and developing strategies for reducing nonresponse is critical to keeping survey costs in check and making survey research efficient. The results of this study suggest that embedding the first question of a web survey in the invitation e-mail may be an effective strategy for increasing response rates. The results of a randomized experiment on over 3300 respondents suggest that the embedded survey question method increased the survey start rate by over 33% (3.5 percentage points) and the final response rate by over 30% (nearly 2 percentage points). These results of this study support prior work demonstrating the influence of design features of the survey invitation on participation. They

also provide empirical evidence of the potential for removing barriers to participation to affect response rates. Both survey start rates and completion rates were dramatically increased by reducing the number of steps that the respondent is required to take between making the decision to participation and responding to the first survey question. These findings will be of value to researchers that employ e-mail invitations to web surveys.

Using a Progressive Engagement Introduction to Gain Cooperation in an Interviewer-administered Survey

Paul J. Lavrakas, *NORC at the University of Chicago*
Jenny Kelly, *NORC at the University of Chicago*
Colleen McClain, *NORC at the University of Chicago*

Gaining cooperation from sampled respondents in interviewer-administered surveys is heavily dependent on interviewers' abilities to avoid immediate hang-ups and counter other forms of refusals. The traditional approach to introducing a survey has the interviewer dominating the exchange with the respondent, because it explicitly and implicitly requires little participation from the respondent beyond stating her/his agreement or disagreement to cooperate. This often leads to too brief contact with the respondent that in turn prevents the interviewer from having time to diagnose how to apply successful persuasion. In contrast, a more successful introduction is hypothesized as promoting a balanced interviewer-respondent exchange that quickly builds rapport with the respondent. This latter approach, which we tested in an experiment, is called Progressive Engagement (cf. Burks et al. 2007), whereby the interviewer sequentially builds rapport with the respondent by progressively engaging the respondent in a back-and-forth exchange about the survey. Our data were gathered in a large dual-frame RDD health survey of California conducted in early 2015 by NORC. The 122K telephone numbers used in the study and interviewers were randomly assigned to either the Convention Introduction or the Progressive Engagement Introduction. The latter was used by interviewers to convey all the same information as in the Conventional Introduction, but the script was structured, and the interviewers were trained, to build in distinctive pauses after each sentence to allow and "invite" the respondent to engage the interviewer with a comment or question. In addition to the experiment, other data were gathered including from recordings of each of the introduction types, monitoring interviewers to determine how well they applied each introduction, and post-survey questionnaire data from interviewers about their use of each introduction. Our results indicate that the Progressive Engagement Introduction was approximately 30% more successful in leading to a completion (p < .001).

"Don't be Afraid … We're Researchers!": The Impact of Informal Contact Language on Response Rates and Patterns of Response

Kelly N. Foster, *East Tennessee State University*
Nicholas E. Hagemeier, *East Tennessee State University*
Arasham A. Alamain, *East Tennessee State University*
Robert Pack, *East Tennessee State University*
Rajkumar J. Sevak, *East Tennessee State University*

Web-based surveys are attractive to researchers for a variety of reasons. The extent to which these results can infer population level characteristics is largely limited by sample design and non-response bias. This project seeks to explore the impact that email contact language has on college students likelihood to complete a survey. Because college students have free access to internet, are usually required to have a university email address, and often receive correspondence through these

email addresses they are an attractive population for internet surveys. Unfortunately, because of this it is more difficult for any one request to stand out enough to get attention. This study expands on findings by Rath et al presented at the AAPOR 2015 conference that contact emails with humor have significantly higer response rates than more traditional formal language. The study includes a random selection of community-college students from all 13 Tennessee Board of Regents Community Colleges (~60,000 students will be included in the study). The larger purpose of the study is to examine non-medical prescription stimulant use among this population. Students will be randomized into one of two invitation conditions – formal and informal language. The initial contact language will be mimicked through all subsequent contacts. Our hypothesis is that students will be drawn to the informal language because it more closely mirrors how they write and speak but we are also interested in determining if the informal language may increase response among prescription drug abusers. Data collection will be completed by the end of March 2014 and the analysis will look at response rates by treatment, patterns in non-response, patterns in rates of use by treatment condition, and demographic differences. This project seeks to add to the body of literature surrounding survey recruitment methods for web-based surveys, particularly for those that involve sensitive questions.

Messaging to Increase Response Rates

Thx 4 Participating!: Evaluating the Effectiveness of Sending Text Messages to Increase Survey Response

Amanda L. Skaff, *Mathematica Policy Research*
Felicia I. Hurwitz, *Mathematica Policy Research*
Jillian Stein, *Mathematica Policy Research*

In a world in which young adults use their phones more for texting than for talking—sending or receiving an average of 87.7 text messages per day—it is unsurprising that telephone interviewers encounter more automated voice messages than people when they try to contact sample members.[1] According to a study by the Pew Research Center, 55 percent of adults who send or receive 51 or more texts per day prefer to be contacted via text message, rather than voice call.[2] In this regard, study members' changing preferences often conflict with traditional approaches to survey research. To address this challenge, successful survey researchers must strive to understand how to adapt and evolve. In an attempt to proactively minimize attrition and maximize response rates, the Evaluation of the YouthBuild Program research team incorporated a random assignment experiment to test whether including text messages in its data collection strategy increased survey response. In this paper, we describe our approach to seeking consent and sending text messages, explain why and how our approach to using text messages has evolved over the lifespan of this longitudinal study, and analyze whether sending text messages encourages survey completion. Though our data collection efforts are ongoing, our research to date suggests that text messaging can play an important part in a multifaceted approach to reaching study members and encouraging survey response. [1] Smith, Aaron. "U.S. Smartphone Use in 2015." Pew Research Center. April 1, 2015. Available: http://www.pewinternet.org/files/2015/03/PI_Smartphones_0401151.pdf [2] Smith, Aaron. "American and Text Messaging." Pew Research Center. September 19, 2011. Available: http://www.pewinternet.org/files/old-media//Files/Reports/2011/Americans%20and%20Text%20Messaging.pdf

The Effect of Using Text Messages for Survey Invitations and Reminders

Kyley McGeeney, *Pew Research Center*
H. Yanna Yan, *University of Michigan*

Text messaging has become ubiquitous but its use in survey research is still limited. It is unclear who can receive text messages as it's a combination of who has a cell phone, has consented to receive text messages and does not have short codes blocked (the six-digit number from which the text is sent). Would using text messages to send invites and reminders to a web survey, in addition to sending emails, increase the response rate or in any other way change response patterns? Does sending text messages change the demographic composition of the resulting sample? To answer these questions The Pew Research Center conducted experiments in two waves of monthly data collection using its nationally representative, probability-based American Trends Panel. In the first experiment, panelists who had consented to receiving text messages from the Panel were randomized to two groups: one received survey invitations via both email and text message while the other half received only email invitations; all reminders were sent only via email. In the second experiment, all panelists who had consented to receiving text messages were sent invites via both text message and email; half of the initial nonrespondents were then randomly assigned to receive reminders via both modes and the other half received only email reminders. We found that those who are able to receive text messages are younger, more highly educated and have higher incomes than those who cannot. In both experiments, the experimental treatment group had a higher click rate and response rate than the control. The treatment groups responded more quickly to the surveys and were more likely to do so using a smartphone. Interestingly, the resulting respondents in the treatment group samples were nearly identical demographically to the control.

Reaching the Mobile Generation: Reducing Web Survey Non-response through SMS Reminders

Kirti N. Kanitkar, *The Gallup Organization*
Jenny Marlar, *The Gallup Organization*

Web Surveys are increasingly being used as alternatives to traditional telephone and mail surveys, for reasons of cost, speed, and presumed appeal to younger demographics. However, web surveys suffer from low response rates, and just as in other modes, younger and minority respondents have lower response rates.Text message (SMS) communication may be an effective alternative to email for improving web survey participation from hard to reach groups. A 2014 Pew report indicated that 90% of Americans own a mobile device and sending and receiving text messages is the most common cellphone activity. Younger people and minorities are more likely to be "mobile dependent -and rely solely on their phone to access the internet. SMS invitations and reminders with links to the web survey may make it easier for mobile dependent respondents to access the survey, thereby improving response rates. SMS reminders have been used effectively outside of survey research, such as in healthcare to improve adherence to treatment and attendance at appointments. However, findings on the use of text messages for survey research has been limited, in large part due to legal restrictions.The present study uses the Gallup Panel to explore the effectiveness of text message reminders for web surveys. Panel members who provided consent to receive text messages were randomly assigned to receive SMS reminders or email reminders. Respondents in the SMS treatment received messages directing them to the web survey. The findings will explore differences in participation rates by reminder mode, as well as by demographics, device used to complete the survey (mobile versus computer), substantive answers, and timeliness of responses. Panelist perceptions of text message reminders compared to email reminders will be discussed.

"When Can We Call?" Experiment to Assess SMS Text to Prompt Response Across Cultures

John Lee Pratt Holmes, *Qatar University (SESRI)*
Abdoulaye Diop, *Qatar University (SESRI)*
Kien Trung Le, *Qatar University (SESRI)*
Anis Miladi, *Qatar University (SESRI)*

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

Effectiveness of Messaging to Encourage Response to the ACS

Jenna Fulton, *U.S. Census Bureau*
Gerson Morales, *U.S. Census Bureau*
Jennifer Hunter Childs, *U.S. Census Bureau*

The American Community Survey (ACS) is an ongoing, mandatory survey conducted by the U.S. Census Bureau that collects essential information about states, local areas, and communities on an annual basis. Data collected in the ACS help to determine how more than $400 billion in federal and state funds are allocated across the nation each year. The ACS samples approximately 295,000 households each month, and responses are collected primarily through self-administered web or mail survey, with non-response follow-up efforts conducted by telephone and in-person interview. However, with declining rates of self-response to this survey (Walker, 2015), there has been an increase in the use of costly follow-up efforts to obtain responses. In an attempt to learn more about how to reverse this trend, we conducted research to identify messages that effectively educate Americans on the importance and legitimacy of the ACS, and encourage self-response to the survey. The U.S. Census Bureau contracted the Gallup Organization to collect survey responses on a nightly basis to assess respondents' stated willingness to participate in the ACS after being presented with a series of messages

intended to convey the relevance of the survey and motivate self-response. The Gallup nightly survey also captured information on respondents' attitudes toward trust, privacy, and confidentiality as they relate to federal statistics, as well as standard demographic information. This research will investigate the relationship between respondents' stated willingness to participate in the ACS given the different motivational messages presented, their attitudes toward trust, privacy, and confidentiality, and demographic information. This will help identify which motivational messages resonate most with respondents to encourage participation, and if the preferred motivational message differs by demographic group or attitudes toward privacy, confidentiality, and trust.

Questionnaire Design

The Effect of Respondent Commitment and Tailored Feedback on Response Quality in an Online Survey
Kristen Cibelli, *University of Michigan*
Frederick G. Conrad, *University of Michigan*

To the extent that inaccurate or poor quality survey responses is due to insufficient effort by respondents, it might help to obtain a direct commitment from the respondent to be diligent in completing the task and to provide feedback "tailored" to the respondent's effort on the response process. Charles Cannell and his associates pioneered these techniques in the late 1970s in face-to-face health-related interviews and the results were promising: The quality of reporting on response tasks designed to be demanding in terms of recall, cognitive effort or self-disclosure increased when they asked respondents to commit to providing complete and accurate information and instructed interviewers to provide feedback on the adequacy of answers (Miller & Cannell, 1977, 1982; Oksenberg, Vinokur, & Cannell, 1977a; Oksenberg et al., 1977b; Vinokur, Oksenberg, & Cannell, 1977). The current study examines the effects of respondent commitment and tailored feedback on the accuracy of reported medical visits in an online survey of parents regarding their child's medical care. The accuracy of responses to several health service utilization questions will be directly validated against medical records from the University of Michigan Health Service – previous studies evaluating commitment and tailored feedback have only used indirect measures of accuracy. The study also examines the effect of respondent commitment and tailored feedback on socially desirable reporting, satisficing, respondent-reported record check, and debriefing questions. It also uniquely assesses the potential benefit of combining these strategies and their effectiveness applied to a web survey. Applying these techniques, which were developed for interviewer-administered interviews, to an online – and therefore interactive – survey promises to promote thoughtful answers in a new generation of survey data collection.

Best Practice Instrument & Communications Evaluation: An Examination of the NSCH Redesign
William Bryan Higgins, *ICF International*
Rikki Welch, *ICF International*
Robert Tortora, *ICF International*
Catherine J. Vladutiu, *U.S. Department of Health & Human Services*

Several national surveys have recently undergone redesign for numerous reasons, including taking advantage of new technologies, improving sampling methods, or including new content areas. The redesign of National Survey of Children's Health and National Survey of Children with Special Health Care Needs falls into all three categories: investigating the use of web data collection for this two-phase

study, introducing an Address Based Sampling design, and combining both surveys into one data collection instrument. This study reviewed the advance communications materials and the redesigned self-administered mail questionnaires. The goal of the communications materials review was twofold: to establish the legitimacy of the survey request, and to evaluate how well the materials encourage potential respondents to complete the four-stage process of opening the material, digesting its content, handing it off to the most knowledgeable adult, and completing the survey. Following the principles identified in Dillman, Smyth and Christian (2014), the review evaluated whether the survey instruments are designed according to sound measurement guidelines, navigable, and have correct skip patterns and accurate question wording. The results provide both an evaluative process for reviewing survey materials and recommendations for changes to the NSCH instruments and advance materials.

Web Survey Experiments on Fully Balanced, Minimally Balanced and Unbalanced Rating Scales
Sarah Cho, *SurveyMonkey*
Mingnan Liu, *SurveyMonkey*

When asking attitudinal questions with dichotomous and mutually exclusive response options, the questions can be presented in one of three ways. A fully balanced question presents both sides of the competing viewpoints. A minimally balanced question presents only one side of the viewpoint but only mentions the competing side very briefly. A third approach is to ask an unbalanced question that only presents one side of the viewpoint while completely ignoring the competing side. Although previous research has compared the fully vs. minimally balanced rating scales, as far as we know, these three types of rating scales have not been tested in a strict experimental setting. Further, previous research has been almost exclusively based on interviewer-assisted telephone surveys, the findings which may not apply to self-administered web surveys. In this study, we report two web survey experiments testing these three types of rating scales among 16 different questions. Overall, the univariate distributions across these three scale balancing types are very similar to one another. Similar patterns are found when breaking down the analysis by respondent's education level. The results suggest that for self-administered web surveys higher efficiency can be achieved through unbalanced dichotomous rating scale questions.

6 Ways to Format the Same Question: But Which One is Correct?
Annie Pettit, *Peanut Labs*

Most researchers have their own question writing style and choose it at every opportunity, even if a different style might be better suited. We sometimes forget that there are many ways to format the exact same question, each one being no more valid than the rest. In this presentation, I will demonstrate six different ways that the same question can be formatted, and how those formats generate different results. These six formats include: 1) Traditional 5 point Likert scale grid 2) Converting a one stage scale into a two stage scale (i.e., Do you agree or disagree? Is your opinion somewhat strong or very strong?) 3) Converting a traditional grid with 5 items into 5 separate questions 4) Using question specific answers (e.g., somewhat happy) rather than generic answers (e.g., somewhat agree) 5) Using the word 'very' as the most extreme label 6) Using the word 'extremely as the most extreme label Although there is likely no true correct answer, it is important to understand how the results differ – does one format create more or less variability, or more or less straightlining? In addition, I will illustrate how various demographic groups respond to these different formats. I will show whether certain demographic groups are more likely to respond with more or less variability or straightlining to a certain format. Is one format easier to understand? Based on these results, researchers will be better able to choose a

question format that creates the type of variability required for their statistical analyses, as well as one that is easily understood and responded to by their target audience.

How Much Can We Ask? Assessing the Effect of Questionnaire Length on Survey Quality

Rebecca Medway, *American Institutes for Research*
Danielle Battle, *American Institutes for Research*
Mengmeng Zhang, *American Institutes for Research*

Survey researchers are often faced with striking a balance between asking all questions to which they would like a response and creating an easier to complete, concise instrument. Existing research suggests that sample members are less likely to respond to longer questionnaires and that responses to items presented later in questionnaires may be of lower quality (e.g., Burchell & Marsh 1992; Galesic & Bosnjack 2009). However, it is difficult to know the 'optimal' length to maximize the response rate and response quality in any particular survey. In this presentation, we will report on the results of a questionnaire-length experiment embedded in the adult training and education component of the mail-based 2014 National Household Education Survey Feasibility Study. This randomized experiment was conducted to determine the optimal questionnaire length for future administrations. Sample members were randomly assigned to one of two questionnaire length conditions. In the first, sample members received a longer questionnaire that included all topics of interest (28 pages, 80 numbered items). In the second, they instead received a shorter questionnaire focusing on a critical subset of these topics (20 pages, 66 numbered items). Preliminary results show that, as expected, the response rate was significantly higher when the shorter booklet was used. This presentation will report on (1) the final response rate in the two conditions; (2) other potential effects of questionnaire length on key survey outcomes, such as sample representativeness, key estimates, and response quality indicators (e.g., item nonresponse, skip errors); (3) whether the use of a prepaid incentive mitigated the impact of questionnaire length; and (4) the implications of these results for future administrations of this and other mail surveys.

Effects a la "Mode"

Examining Trends in the Presence of Survey Mode Effects

Heather Hisako Kitada, *Oregon State University*
Virginia Lesser, *Oregon State University*

Due to differences in cognitive and social stimuli affecting how an individual completes a questionnaire, survey researchers must deal with the consequences of mode administration differences. Mode choice affects data quality, measurement error, and estimation, potentially having implications on inference. Modes of data collection change due to lower cost, convenience, and representativeness of a sampling frame. Data collected from repeated surveys conducted over time may have incorporated different data collection modes. These changes are often associated with an inability to detect changing opinions due to possible confounding bias of mode effect. From 2006 to 2014, a biennial survey has been conducted of Oregon residents assessing satisfaction with highways and determining the most important issues facing Oregon transportation. A series of experiments were conducted on the general population to determine the impact of mode on survey results and to examine the demographics of respondents across each of these modes over time. All surveys were based on a stratified random sample of Oregon residents. For all years, mail and mixed-mode (Web followed by mail) were used to

compare results. The Oregon household contact information used address-based sampling with addresses obtained from the US Delivery Sequence File. In 2006 and 2008, a random digit dialing telephone mode was also used. We implement a design based modeling perspective to estimate mode effects through time after accounting for demographics covariates such as gender, age, income, education attainment, and race. We also present a simulation study to assess the sensitivity of our proposed bias correction methodology.

Household Participation and Response Quality in a Survey Featuring Mobile and Desktop Modes

Brett McBride, *U.S. Bureau of Labor Statistics*
Ian Elkin, *U.S. Bureau of Labor Statistics*
Lucilla Tan, *U.S. Bureau of Labor Statistics*

Studies have examined the benefits of offering prospective respondents to mixed-mode surveys their preferred mode option (Olson et al., 2012). However there are cost and operational reasons why a survey might request that individuals use a specific mode – via either a sequential mixed-mode design or as a single-mode design. These reasons were in effect in the Consumer Expenditure Diary Survey's test of a mobile-optimized web diary in which household members recorded their daily expenses. The study design entailed interviewers asking eligible household members to record expenses in a diary using a mobile-optimized web mode (if they possessed a mobile phone). If they did not possess a mobile phone, they were asked to record expenses using a desktop-optimized web mode, if they had internet access. Household members without internet access were not eligible for the test. This single-mode design (from the respondent's perspective) resulted in a variety of within-household mode combinations. This research examined three of those combinations – mobile-only households, desktop-only households and households having a mix of modes. The research describes how these three groups differed on within-household survey participation and data quality, and presents some hypotheses for the differences that were observed. Findings from the research suggest the benefits to providing household members a choice of options for online survey response.

Mode and Recruitment Effects on Item Non-response of Worksite Information Reported by Nurse Practitioners

Michelle L. Cook, *American Association of Nurse Practitioners*
Lydia E. Riley, *American Association of Nurse Practitioners*
Tyler Litsch, *American Association of Nurse Practitioners*

The American Association of Nurse Practitioners (AANP) has been conducting the National Nurse Practitioner (NP) Practice Site Census periodically since 2002 in order to collect current and essential data as they relate to NP practice locations and characteristics. Historically, data have been collected bi-modally, by internet and mail. In the fall of 2015, AANP piloted the survey tool for the 2016 Census with a group of 1,197 nurse practitioners. The sample was divided into eight experimental groups which varied by recruitment strategy (mail, email, and/or SMS text message) and by mode (online and/or mail survey). Item non-response on worksites were analyzed and compared between the different recruitment types and by survey mode to determine whether or not recruitment strategy or survey mode impacted the reporting of this important information. Online survey respondents (through either SMS or email invitations) were just as likely to fill out their worksite information as those who participated by mail. However, participants who responded via email invitation had the lowest percentage of item non-response on worksite address fields, in particular Company Name (7.0%), Street

Address Line 1 (14.0%), and Zip Code (12.5%) which were significant at the 0.10 level. Those who responded through an SMS message or mail invitation had very similar rates of item non-response on all address fields. Additional questions tied to each worksite (e.g., number of hours worked at that site, type of position held, etc.) also showed that those recruited by text message were more apt to not respond. Introducing SMS message invitations may be a viable option to increase overall response rates, but could introduce bias into the reporting of results.

The Impact of Response Scale Direction on Survey Responses in a Mixed-mode Survey

Mengyao Hu, *University of Michigan*
Ting Yan, *Westat*
Florian Keusch, *University of Mannheim*

So far, no conclusive empirical evidence exists that would help guide questionnaire designers to make informed decisions on whether rating scales should ascend from the positive pole (or highest level) to the negative pole (or lowest level) or should descend from the negative pole (or lowest level) to the positive pole (or highest level). While previous research found mean shifts and changes in proportions endorsing various response options under different scale directions, but little is known about the influence of scale direction on other properties of the resulting answers (e.g., factor structures, internal consistency of resulted factors). Also, only limited attempts have been made in trying to reveal the underlying cognitive mechanisms at play when answering response scales. In an experiment in the UK Innovation Panel administered face-to-face and over the Web, respondents were randomly assigned to answer two sets of multi-item batteries (5 items on life satisfaction and 20 items from the Balanced Inventory of Desirable Responding) on a descending scale or an ascending scale. Preliminary analyses indicate that the scale direction significantly affects resultant survey responses by pushing answers to the beginning of the scale. We found that the mode of data collection does not interact with scale direction – scale direction effect is shown in both the CAPI and the Web mode. Further analysis will focus on the influence of scale direction on factors structure and internal consistency of the resulted factors and the role of satisficing in explaining response behavior to rating questions.

Mode and Eligibility Rates in a Dual-mode Web and Mail Survey

Ilana Ventura, *NORC at the University of Chicago*
Ipek Bilgen, *NORC at the University of Chicago*
Michael J. Stern, *NORC at the University of Chicago*

Mode effects have been studied extensively with regard to response rates (Shih and Fan, 2008), disclosure of sensitive information (Kreuter et al 2008; Atkeson et al 2014), data quality (deRada and Domínguez-Álvarez, 2014), and eligibility (Tourangeau, 2012), but usually within the context of comparing interviewer mediated and self-administered modes. This paper expands on extant work by examining the relationship between mode, eligibility, and response rates, for two self-administered modes, Web and mail. Our primary research question is whether mode has an influence on screener completion rates. In addition, we examine the interaction between mode effect and question sensitivity on respondents completing the questionnaire. We use data from a study on U.S. residents' health behaviors, opinions on family structure and roles, relationships, and sexual behavior and identity. A national probability-based address-based sample (ABS) of 2,000 households was randomly split into two groups. The first group received a recruitment letter, with a URL and a PIN, instructing the respondent to

complete the survey on the Web, along with a $2 pre-incentive. The second group received a recruitment letter with a questionnaire booklet, along with a $2 pre-incentive. Both the Web and mail groups were then randomly divided into two modules: 1) No Screener– survey all respondents 2) Screener– survey only respondents with children. The screener identified households with children. Screener questions were also asked in the non-screener group within the questionnaire. Preliminary results indicate that the Web performed slightly better than mail, with respect to a higher proportion of returned screeners for families with children, however, the mail mode had an overall higher response rate. These findings will contribute to the literature on the usage and impact of using screeners in multi-mode surveys.

Advances in Measuring Sexual Minorities (LGBTQ) and Same-sex Households and Relationships

Panel organizer: Matt Jans, *UCLA Center for Health Policy Research*

Tremendous change in the U.S. social and political environment has brought mainstream attention to sexual minorities. Major surveys have begun or expanded their measurement of sexual orientation and gender identity to meet the growing demand for data on these populations. Given this changing survey environment, it is more important than ever to evaluate and implement methods of measuring gender identity, sexual orientation, and same-sex household composition. The major challenge is to do this in a manner that adequately and accurately represents all respondents' identities and also provides useful statistics on these populations. To these ends, this panel will discuss current and developing research on measurement of sexual/gender identity, sexual orientation, and relationship status for same-sex couples. Panelists will discuss findings from national and state surveys and survey experiments. Papers will focus on multiple sources of survey error, addressing practical measurement concerns in order to clear confusion around the measurement of sexual/gender identity and sexual orientation and offer best practices, as well as highlight pitfalls to avoid. Focus will be given to promoting future research on these topics, as well as reporting on current findings.

Identifying Sexual Minorities in Surveys: Research Synthesis and Meta-analysis
Justine Bulgar-Medina, *University of Massachusetts Boston*

Numerous surveys and studies have attempted to estimate the proportion of the United States population that identify as sexual minorities (lesbian, gay, bisexual or transgender (LGBT) via direct data collection or analyses of secondary data (Black et al 2000; Gates 2014). Accurate estimates of member of various social identities have long been utilized in better understanding issues ranging from fair housing to education to employment but often do not include (at all or accurately) measures of sexual orientation, as they do for gender or race and ethnicity (Cameron et al 2009). When measures of sexual orientation identity are included directly (e.g. not via proxy using questions of sexual behavior), the exact wording and mode of survey administration varies producing slightly different estimates. The present work synthesizes the existing body of knowledge by considering numerous studies and sources of secondary data. Published studies for inclusion were identified in social science journals via WorldCat and unpublished sources were compiled using cross-references and targeting key sources (e.g. the Williams Institute at UCLA). By including both large-scale national datasets and other data sources, this work considers a broad array of literature and data sources, including large-scale national surveys such as the National Survey of Family Growth, US Census, and the General Social Survey, as well as smaller and more localized studies. Ultimately, the goal of this work is to offer additional estimates of sexual

minorities in the United States population, better understand methods of obtaining this information more accurately in survey and potentially expanding by offering estimates of multiple minority identities.

Testing Web-Based Survey Measures of Gender Identity and Sexual Orientation Using Mark-All-That-Apply and Forced Choice Questions

Philip S. Brenner, *University of Massachusetts Boston*
Justine Bulgar-Medina, *University of Massachusetts Boston*

Best practices in measuring sexual orientation and gender identity suggest exhaustive but not necessarily exclusive response options. This goal implies the use of a mark-all-that-apply (MATA) to adequately measure these identities. However, problematic features of MATA can harm data quality. Consequently, forced choice — changing each response into a yes/no question — is typically preferred. While potentially increasing response validity, forced choice also increases respondent burden, especially on smartphones and for longer lists of response options. Thus, the trade-off between forced choice and mark-all-that-apply is between respondent burden on the one hand (which leads to increased breakoffs, partial interviews and unit nonresponse) and the potential for measurement error, on the other. But what if the topic is of high salience for the respondent? Highly motivated respondents are more likely to agree to participate and give higher quality responses, potentially overcoming some design problems. Thus, our research question is this: can respondents for whom the questions are salient overcome the flaws of the MATA? Building on pilot work, this survey experiment randomly assigns respondents from a web panel (with an LGBT oversample) into categories in a 2x3 factorial design. The first factor is question type: (1) MATA or (2) forced choice. The second factor is question presentation: (1) sexual orientation and gender identity in one question, (2) in two questions on one page; or (2) in two questions on two pages. Findings contribute to the body of knowledge on the self-identification practices of this specific population as well as methodological considerations emerging from satisficing in topics of high salience.

Improving Measurement of Same-sex Couple Households for Census 2020

Rose Kreider, *U.S. Census Bureau*
Daphne Lofquist, *U.S. Census Bureau*
Nancy Bates, *U.S. Census Bureau*
Jamie Lewis, *U.S. Census Bureau*

The Census Bureau has been working to improve measurement of same-sex couple households, especially for same-sex married couples. In 2015, several tests of the revised relationship question were conducted. The new question lists explicit categories for "opposite-sex husband/wife/spouse," "same-sex husband/wife/spouse," "opposite-sex unmarried partner," and "same-sex unmarried partner." This presentation will provide an update on those testing activities. A relatively small population such as same-sex married couple households are difficult to count since it only requires a very small percentage of opposite-sex couples to mismark relationship or sex in order to affect the estimate of same-sex married couple households. In anticipation of respondent mismarks, an automated check was added to the self-response internet and mobile device nonresponse follow-up (NRFU) instruments. The presentation will evaluate the efficacy of these checks. Tests in the spring of 2015 included the new relationship question across three modes. The purpose of the tests was to examine NRFU operations and test various methods to increase self-response. The larger National Content Test (NCT) was a self-response test (internet and mailed questionnaires) conducted in the fall of 2015. The NCT included both

a test panel with the new relationship question, and a control panel with the former relationship question. This was the largest test to date, comprising about 1.2 million households. The internet mode included the consistency check between relationship and sex reports for coupled households. Using response data and paradata, the paper will report on item nonresponse rates, consistency of reports between sex and relationship, and degree of misreporting and distribution of categories between the test and control versions.

Measuring Gender and Sexual Identity in a Nationally Representative Survey of Older Adults: Results from Qualitative Research using Cognitive Interviews

Stuart Michaels, *NORC at the University of Chicago*
Michael J. Stern, *NORC at the University of Chicago*
Michael Kozloski, *NORC at the University of Chicago*
Paul Guerino, *U.S. Centers for Medicare & Medicaid Services*
Samuel C. Haffer, *Data and Policy Analytics Group*
Debra Reed-Gillette, *U.S. Centers for Medicare & Medicaid Services*

There is a significant need to improve measures for sexual and gender identity in representative population surveys in order to address health and other disparities experienced by sexual and gender minority groups. Even though there has been great progress in developing measure of sexual identity (by NCHS for the NHIS) and an increase in the number of representative national surveys that include sexuality items, our knowledge and experience is limited about the comprehension of these items by older adults. In addition, we are only just beginning to develop and use items to distinguish the transgender population. In order to address this gap in the research, we report on a series of cognitive interviews with Medicare beneficiaries focusing on older adults, English and Spanish speakers, LGBT and non-LGBT respondents. After a review of existing items, the performance of these items in population surveys, and the selection of items for testing, we present the results of cognitive interviews to test the performance of a sexual identity question and a two-step approach to measure gender identity for possible inclusion in a survey of Medicare beneficiaries. Sixty cognitive interviews were conducted with a diverse group of Medicare beneficiaries (most over 65 years old) with attention paid to Spanish speakers and both LGBT and non-LGBT subjects.

Data Quality Research of LGB Measures at the National Center for Health Statistics

Kristen Miller, *National Center for Health Statistics*

In 2013, the National Center for Health Statistics developed a sexual identity question for the National Health Interview Survey (NHIS) with the goal of improving sexual orientation measurement in nation-wide population-based surveys. Resulting NHIS data reveal that the new question produces lower rates of missing data than any other comparable survey. Since this development work, NCHS continues to examine aspects of LGB data quality, including interview and mode effects, survey context and question placement. The presentation will summarize ongoing activities at NCHS and discuss directions for future research.

Measuring Initial Results of Gender Identity Questions in the 2015-2016 California Health Interview Survey

David Grant, *University of California, Los Angeles*
Matt Jans, *UCLA Center for Health Policy Research*
Royce Park, *University of California, Los Angeles*
Ninez A. Ponce, *University of California, Los Angeles*
Jody L. Herman, *The Williams Institute*
Bianca D. M. Wilson, *The Williams Institute*
Gary Gates, *The Williams Institute*

A large population-based health survey of the nation's most diverse state, the 2015-2016 California Health Interview Survey (CHIS) included new gender identity (GI) questions to measure transgender adults. Following a pilot-test of 4 different versions conducted in 2014, a "two-step" GI measure was asked of the full adult CHIS 2015-2016 sample since May 2015. Evaluation of the 2014 GI pilot-test led to two modifications implemented in CHIS 2015-2016. First, the first gender identity question wording was modified to include the response options (male or female) in the question itself because interview monitoring revealed frequent pauses among respondents who seemed uncertain as to an appropriate response. The question was revised from, "What sex were you assigned at birth, on your original birth certificate?" to "On your original birth certificate, was your sex assigned as male or female?" Because the revised wording was not tested before fielding, adult respondents were randomly assigned to either the original or modified wording. The second modification to CHIS 2015-2016 GI administration was adding a confirmation question for respondents who reported a change in gender identify from assigned sex at birth to inform potential measurement error. Overall item nonresponse to the GI questions during the 2014 pilot was very low (generally less than 2%). However, the item nonresponse rate was approximately six-times higher than the proportion of adults who identified as transgender. This makes gender identity particularly vulnerable to measurement error due to respondent error or confusion, data entry error by interviewers, or other question administration errors. To minimize such errors at this early stage of GI measurement in CHIS, the confirmation question was added to verify positive transgender responses. This presentation provides results on performance of the GI questions and modifications for the initial approximately 10,000 adult CHIS 2015-2016 respondents.

Nonprobability Sampling: Fast, Cheap . . . And Accurate?

Non-Representative Surveys: Fast, Cheap and Mostly Accurate

David Rothschild, *Microsoft Research*
Sharad Goel, *Stanford University*

Probability-based sampling methods, such as random-digit dialing (RDD), are a staple of modern polling and have been successfully used to gauge public opinion for nearly a century. Though historically effective, these traditional survey techniques are often slow and expensive. At the same time, it has become increasingly quick and cost-effective to collect non-probability-based convenience samples. Here we investigate the potential of such non-representative data for survey research by administering an online, fully opt-in poll of social and political attitudes. Our survey consisted of 49 multiple-choice attitudinal questions drawn from the 2012 General Social Survey (GSS) and select surveys by the Pew Research Center. To correct for the inherent biases of non-representative data, we statistically adjust estimates via model-based poststratification. Compared to typical RDD polls, the online survey required

less than one-tenth the time and money to conduct. After statistical correction, we find the median absolute difference between the online survey and the gold-standard GSS and Pew studies is 7.4 percentage points. Though this difference is considerably larger than if the surveys were all perfect simple random samples, we find the gap is comparable to that between the GSS and Pew estimates themselves, ostensibly because even the best available surveys suffer from substantial non-sampling error. We conclude that non-representative surveys are a promising tool for fast, cheap, and mostly accurate measurement of public opinion.

Polling: Cost, Speed and Accuracy

Solomon Messing, *Pew Research Center*
David Rothschild, *Microsoft Research*
Tobias B. Konitzer, *Stanford University*
Sharad Goel, *Stanford University*
Courtney Kennedy, *Pew Research Center*

Survey data can be gathered along a spectrum of quality from pure opt-in samples with limited coverage to perfect probability samples with perfect coverage. As the quality of the data collection increases, so does the cost in both time and money. This paper gathers and examines 10 data collection methods across as many as 60 different questions. We then run the data through several different analytics packages, from raking to modeling and post-stratification.The paper does not pick winning and losing data collection or analytics methods, but describes the trade-offs. For some outcome spaces the accuracy provided by the lower quality data is off-set by the cost (especially true if the outcome space is more about movement than levels); for others it is worth the expense to provide small, but statistically significant gains in accuracy. Similarly, the right analytics packages is not universal, but the value shifts by both data type and question time. On the extreme end is multilevel regression post-stratification (MRP). Work exists showing that MRP can help adjust for even extreme selection biases in sample of Americans (e.g., Xbox users, Wang et al., 2013) when estimating vote choice with extreme opt-in sample with limited coverage. While we confirm its value for some data and question type combinations, raking is more universal and less costly, and converge with MRP for some other data and question type combinations. This paper illustrate the trade-off with the accuracy of the final population estimates of sentiment as we increase the quality and cost of data collection and the analytics used to transom that raw data into population estimates of sentiment.

The Utility of Non-probability Samples for Assessing Population-level Health Information Seeking Behavior

Reanne L.M. Townsend, *Westat*
Hanyu Sun, *Westat*
Aaron Maitland, *Westat*
Terisa Davis, *Westat*
J. Michael Brick, *Westat*
Richard P. Moser, *National Cancer Institute*
Gordon Willis, *National Cancer Institute*
Kelly Blake, *National Cancer Institute*
Stephanie Fowler, *National Cancer Institute*

The goal of this paper is to help to inform the literature on non-probability sampling by comparing estimates produced from three non-probability surveys to estimates from a national

probability sample. Surveys utilizing probability samples of a target population are widely used in a variety of research disciplines. However, survey participation rates have been declining for several years, leading to increased cost to recruit sample members and concerns about the representativeness of the data collected. As the amount of effort required to make valid inferences from probability samples has increased, non-probability samples have become an increasingly used alternative. Although there have been a variety of previous comparisons between non-probability and probability samples, the field of survey research is still struggling to understand the circumstances under which non-probability samples yield valid inferences of population estimates. Further, there is no single approach to non-probability sampling, as this general term subsumes a range of specific practices that may influence the cost, representativeness, and usefulness of the collected information. We will examine estimates produced from these three non-probability samples which represent alternative approaches to recruiting sample members, and vary in the cost of obtaining the sample. These estimates will be compared with a benchmark national probability sample survey of households about general health and health information seeking behavior. We will examine differences with a variety of statistics including point estimates, subgroup differences and associations between variables. Finally, we will suggest conclusions concerning the trade-offs involved in selecting fundamentally different approaches to the collection of self-report information through non-probability sampling.

What Explains Variability in Survey Estimates from Nonprobability Samples?

Scott Keeter, *Pew Research Center*
Andrew Mercer, *Pew Research Center*
Courtney Kennedy, *Pew Research Center*
Kyley McGeeney, *Pew Research Center*
Nicholas Hatley, *Pew Research Center*

In its 2013 report, the AAPOR Task Force on Nonprobability Sampling wrote that "Unlike probability sampling, there is no single framework that adequately encompasses all of non-probability sampling. Non-probability sampling is a collection of methods and it is difficult if not impossible to ascribe properties that apply to all non-probability sampling methodologies." Because of this, it is not surprising that across 10 nonprobability samples from 10 different providers, the median item-level difference in 60 survey estimates was 4 percentage points, while the range of differences on individual items varied between 4 and 34 points. In this paper we report on a large scale study of data collected from 10 different nonprobability providers using a common questionnaire with 60 frequently used survey measures across a wide range of topics. The paper will describe the pattern of results, including the variability of the estimates, how closely they match estimates from probability sample surveys and government benchmarks and how much traditional weighting adjustments such as raking reduce the variability of estimates. In an effort to understand why estimates differ, the analysis will also examine the relationship between sampling and data collection features of individual organizations, the demographic composition of the samples and the resulting survey estimates. The analysis will attempt to test the assertion of the AAPOR Task Force that "The most promising non-probability methods for surveys are those that are based on models that attempt to deal with challenges to inference in both the sampling and estimation stages."

Comparing Nonprobability Sample Estimates to Probability Sample Estimates

Ronaldo Iachan, *ICF International*
John M. Boyle, *ICF International*
Lew Berman, *ICF International*

This paper presents results of a non-probability sample selected and fielded with a more rigorous data collection protocol. The study extends a previous study (Boyle et al., 2015; Iachan et al., 2015) that demonstrated the comparability for three metropolitan areas. That initial study adopted the protocol typically followed in internet panels with no reminders or follow-ups. In addition, that approach controlled the distribution of respondents by selectively adding samples. By contrast, the new study in Los Angeles County included an arm using a more rigorous protocol with several follow-ups and reminders, as well as an arm where that same original protocol was preserved for comparison purposes. For the new arm, in addition to the more rigorous protocol, balancing to the Census demographic distribution in LA County was implemented in the initial sampling and in the post-stratification weight adjustments (raking). We first compare the unweighted distributions of the two arms with the 2014 American Community Survey (ACS) data for LA County. We then describe the weighting procedures, and compare the weighted distribution of the two arms with a number of probability sample surveys. First, we compare with surveys which represent the same area (LA County) and include similar survey questions: a) the SMART Behavioral Risk Factor Surveillance System (BRFSS), and b) the LA subset of the California Health Interview Survey (CHIS). In addition, we compare with national norms for the same questions estimated from the National Health Interview Survey (NHIS) and the Health Information National Trends Survey (HINTS), cycle 4.

Strategies for Mobile Data Collection in International Settings, sponsored by WAPOR

Evaluating a Modular Design Approach to Collecting Survey Data Using Text Messages

Brady T. West, *University of Michigan*
Dirgha Ghimire, *University of Michigan*
William G. Axinn, *University of Michigan*

We present analyses of data from a pilot study in Nepal that was designed to provide an initial examination of the errors and costs associated with an innovative methodology for survey data collection. We embedded a randomized experiment within a long-standing panel survey, collecting data on a small number of items with varying sensitivity from a probability sample of 450 young Nepalese adults. Survey items ranged from simple demographics to indicators of substance abuse and mental health problems. Sampled adults were randomly assigned to one of three different modes of data collection: 1) a standard one-time telephone interview, 2) a "single sitting" back-and-forth interview with an interviewer using text messaging, and 3) an interview using text messages within a modular design framework (which generally involves breaking the survey response task into distinct parts over a short period of time). Respondents in the modular group were asked to respond (via text message exchanges with an interviewer) to only one question on a given day, rather than complete the entire survey. Both bivariate and multivariate analyses demonstrate that the two text messaging modes increased the probability of disclosing sensitive information relative to the telephone mode, and that respondents in the modular design group, while responding less frequently, found the survey to be significantly easier. Further, those who responded in the modular group were not unique in terms of

available covariates, suggesting that the reduced item response rates only introduced limited nonresponse bias. We will conclude by discussing important future research directions aimed at 1) enhancing this methodology, 2) applying it with other modes of data collection (e. g., web surveys), and 3) continuously evaluating its effectiveness from a total survey error perspective.

Adapting Surveys to the Mobile World: Data Chunking in the Dutch Probability-based LISS Panel
Vera Toepoel, *Utrecht University*
Peter Lugtig, *Utrecht University*

Mobile phones are replacing key tasks formally done on PC and laptop. Mobile phones are commonly used for short messaging.We should therefore 1)   move to mobile or mobile-friendly surveys; 2)   find ways to shorten questionnaires (most surveys are too long for mobiles). A way to reduce survey length is data chunking: offering questionnaires in smaller pieces. Data chunking a.k.a. modular survey design is a way to cut-down long survey questionnaires. While data chunking is not new to the survey world (see Johnson, Kelly & Stevens, 2011), there exists no systematic research in how data chunking is related to Total Survey Error and mobile surveys. There are several ways to do this: 'across respondent' modularisation, whereby different respondents take each piece, and 'within respondent' modularisation, whereby the same respondent is permitted to take pieces of a survey at different times. A longitudinal survey in the probability-based LISS Panel of CentERdata is divided in several experimental groups to investigate data chunking within respondents: panellists who own a mobile phone with Internet connection are randomly assigned to a) normal survey length, b) survey cut in 5 pieces, c) survey cut in 10 pieces. In addition, we experiment with (push) notifications via mail and sms. We investigate the number of complete and incomplete responses and look at indicators for data quality (straightlining, primacy, survey length, satisfaction with the survey etc.). In addition, we use multiple imputation to investigate if the imputation could increase data quality to a level equal to the full survey. We benchmark our data with other panel variables, including previous waves of the core questionnaire we used. This paper is highly relevant in times were response rates are falling and survey length should be kept to a minimum.

Social Media Use and Online Civic and Political Participation in China
Weiwu Zhang, *Texas Tech University*
Yunjuan Luo, *Texas Tech University*

Political communication scholars have primarily examined the role of social media use in civic and political engagement in the U.S. context and lack of scholarly attention has been paid to the social and political implications of social media use in non-US contexts. This study attempts to fill this gap by investigating how social media use influences online civic and political participation in China, a transitional society with the largest Internet population in the world. Through a national online survey (N=1,017) conducted in China in August 2015, this study attempts to examine the role of social media use among Chinese netizens in online civic and political participation after controlling for the influences of demographic variables and political variables. Preliminary descriptive findings indicate that social media are widely used among Chinese netizens, in particular, among younger people. In addition, Chinese netizens more heavily rely on social media for informational purposes rather than relational and recreational purposes. Despite a high level of social media use, Chinese netizens are not very active in civic and political engagement online. Political efficacy, willingness to speak up, and reliance on traditional media for information are found to be positively related to both online civic and political participation. After controlling for all the demographic and political variables, reliance on social media

for relational and recreational purposes positively predicts both online civic and political participation. Reliance on social media for informational purposes is found to be a significant predictor of online political participation, but not online civic engagement, suggesting the important role of social media as an information source in enhancing political participation online. Implications of the findings for social media engagement and democratic governance are discussed.

Unintentional Mobile Respondents in Official Statistics and Their Effect on Data Quality
Jeldrik Bakker, *Statistics Netherlands*

An increasing proportion of web surveys are completed on tablets and smartphones, even if the questionnaires are not designed for mobile completion. Earlier research has shown that differences in data quality between answers completed on different devices are often small. Most of these studies have however relied on experimentation on volunteer samples to compare differences in data quality across different devices. In this study, we use rich register data available from the sampling frame available to Statistics Netherlands, to look at data quality for general population surveys. Specifically, we will look at both selection and measurement effects across different devices for a number of Statistics Netherlands' social surveys. In this paper, we compare desktop/laptop, tablet and smartphone respondents in terms of background variables, unit and item non-response, and a range of data quality indicators like percentage breakoffs, average completion times, question order effects, social desirability, straightlining, the length and rounding of open answers and percentage don't know answers. We give recommendations for optimization of the designs over devices.

The Mobile Web Only Population: Socio-demographic Characteristics and Potential Bias
Marek Fuchs, *Darmstadt University of Technology*
Anke Metzler, *Darmstadt University of Technology*

In recent years the use of mobile devices such as smart phones or tablet PCs to complete Web surveys has grown steadily. Thus, survey researchers are facing new challenges when designing Web survey questionnaires. Among others, screen sizes are considerably smaller for mobile devices as compared to desktop PC or notebooks. Also, mobile Web respondents are using touch screen data entry instead of a mouse. Accordingly, the question arises whether Web survey questionnaires should be optimized for mobile devices in order to avoid underrepresentation of respondents using mobile devices due to difficulty arising from the small screen size and touch screen technology. Since most Web survey respondents still have a desktop PC or notebook available to access the internet optimizing questionnaires for mobile devices may not be necessary avoiding potential differential measurement error. In this paper we assess the so-called mobile Web only respondents who no longer have access to the internet using a desktop PC or notebook. It is assumed that this group is most likely underrepresented in Web surveys not optimized for mobile devises resulting in a potential bias. Using Eurobarometer data from 2012 to 2014 across 27 European countries we estimated the size of the mobile Web only population. Results indicate that the percentage of mobile Web onlys increased from 1.5 percent in 2012 to 6.2 percent in 2014 across Europe. A more detailed analysis indicated that the mobile Web only population is younger, more likely female, less educated, more often living in small or middle sized towns and more often single. Overall, results suggest that Web survey questionnaires should be optimized for mobile devices since exclusion or underrepresentation of the mobile Web onlys may cause considerable bias.

Unpacking Apathy and Cynicism in the United States

Are Nonvoters Dissatisfied or Just Disengaged?

Lunna Abrantes Lopes, *Public Policy Institute of California*
Eric McGhee, *Public Policy Institute of California*

The question of how to best convert nonvoters into voters is one that continues to plague civic organizations and political campaigns. With negative coverage of government institutions highlighting polarization, inaction, mismanagement and at times corruption, are nonvoters just too dissatisfied with government to participate in the process? Or are they happy with government but just apathetic about getting involved? Using data from statewide surveys conducted by the Public Policy Institute of California (PPIC), this paper will examine how California's nonvoters differ from voters in their attitudes towards government. Using a combination of survey questions and statistical matching to the voter file, we first identify voters and nonvoters and then compare their responses on a wide battery of questions about trust in government and interest in politics. By utilizing PPIC data we will be able to analyze different subgroups of nonvoters (e.g. age, education, income, immigrant status, race/ethnicity) and compare these groups to their voting counterparts. The results suggest that nonvoters are actually somewhat more trusting of government, and support far more government action in all phases of life. But they are also far less likely to express interest in politics or to feel that their participation will make much difference.

Courts, Campaigns and Corruptions: An Empirical Test of the "Appearance of Corruption" Rationale for Campaign Finance Regulations

Nitya Rao, *The University of Texas at Austin*

There is a broad debate between scholars of American politics over the necessity of individual campaign contribution limits. In Buckley v. Valeo, the Supreme Court argued that these limits were necessary to prevent not only actual corruption, but also the mere appearance of corruption. Specifically, the Court contended that the appearance of corruption alone might undermine the health of a representative democracy. There has been little empirical evidence to support this assertion, so to test this hypothesis, I used regression analysis to determine the effect of perceptions of corruption on two different indicators of democratic health. In the first model, using a question asking citizens whether they believed the country was moving in the right direction as a proxy for democratic health, I found that perceptions of corruption had no effect on democratic health. In the second model, using a political participation scale as a proxy for democratic health, I found that perceptions of corruption had a significant, but positive, effect on political participation. Thus this second analysis also failed to lend support to the hypothesis set forth by the appearance of corruption rationale. In pursuit of a more robust research design, I fielded a novel survey containing different measures of factors influencing perceptions of corruption, perceptions about campaign contributions, support for campaign finance reform initiatives, perceptions of the frequency and nature of corruption, and perceptions of democratic health. I then constructed a structural equation model with the novel survey data to test sequential hypotheses derived from the specific argument contained in Buckley v. Valeo and to determine the stability of various measures of attitudes concerning political corruption.

Understanding Public Cynicism Toward the Federal Government: What is it About Government that Drives Overall Views?

Alec Tyson, *Pew Research Center*
Jocelyn Kiley, *Pew Research Center*
Meredith Dost, *Pew Research Center*

By almost every measure, Americans hold negative overall assessments of the federal government. Today, about three-quarters of Americans say the federal government is doing an only fair or poor job running its programs, only about two-in-ten say they can trust the government in Washington to do what is right most of the time, and most say they are either frustrated or angry with the government. This paper will investigate the relationship between these general assessments of government, and views of several aspects of government. In particular, how do ratings on these overall measures relate to views of how well the government does its job (its performance), views about the preferred scope of government (its role), and perceptions of the political actors and the political system? The principle data source for this investigation is a fall 2015 Pew Research Center survey of over 6,000 respondents, The survey contains several measures of overall opinion of the federal government, as well as views about different aspects of government and politics, including measures about political institutions, elected officials, as well as both broad and specific assessments of both the government's performance in thirteen key areas and the extent to which people think the government should play a role in each of these areas. Additional measures in the survey, including measures of political participation, allow us to examine how views outside the realm of government are associated with feelings about government.

Beyond Right Track/Wrong Direction: A New Multi-indicator Measure of Political Sentiment

Donald P. Levy, *Siena College Research Institute*
David Redlawsk, *Rutgers University*
Ashley A. Koning, *Rutgers University*
Harry Wilson, *Roanoke College*
Meghann Crawford, *Siena College Research Institute*

Many surveys begin asking respondents whether they think the nation/state is on the right track or headed in the wrong direction. Pollsters use this question to describe the mood of the sample or its current assessment of the state of country. Right track is tracked over time, between groups and is analyzed relative to events or ongoing dynamics. This paper presents a new multi-indicator measure of political sentiment and chronicles its development, testing, validation and initial findings as a collaborative project conducted by three academically situated polling centers. While political sentiment as measured by right track/wrong direction can only have two dimensions, we argue that voters are routinely buffeted by multiple foreign and domestic currents. It goes without saying that any one respondent or group of respondents may feel that the nation or their state is on the right track in some arenas and headed in the wrong direction in others. This multi-indicator measure provides a sentiment score for each identified factor cluster including international issues, domestic concerns as well as issues of security and progress and a composite score that considers all the questions that impact the collective mood. The composite score is compared to the well-established right track/wrong direction. The initial instrument was developed through the collaborative efforts of the three centers and tested through data gathered through a national online panel. After submitting the data to factor analysis, trimming the instrument and assessing the reliability of the sub-dimensions, each center

conducted a statewide poll of approximately 800 completes in their home area – New York, Virginia and New Jersey. Future applications of this measure include a new more nuanced tracking mechanism over time as well as a more substantive explanation of how and which events move the mood of the electorate as a whole or in part.

How Do Americans Want Their Elected Representatives to Make Laws?
Divergence in Preference and Perception Drives Disapproval of Congress

Bo MacInnis, *Stanford University*
Sarah Anderson, *University of California, Santa Barbara*
Jon A. Krosnick, *Stanford University*

The American public has disapproved of the U.S. Congress' performance for many years, but the cause of this disapproval is unclear. The public could be unhappy about the gridlock in Congress. The public could be unhappy about the content or quality of the laws Congress has passed. Or the public could be unhappy about another criterion that this project introduces for the first time: the way that Congressional Representatives have been making decisions about how to vote. The process of Congressional decision-making has been much debated from a normative perspective, so it is interesting to use survey methods to explore not only what criteria Americans want Congressional representatives to use when making decisions on how to vote but also what criteria people think the representatives have actually been using. A survey of a nationally representative sample of American adults reveals that Americans want their representatives to pay substantial attention to the general public and to the issue public (people who feel strongly about the issue and attach high personal importance to the issue). On the other hand, the public perceives that their representatives fail to pay attention to the exactly these segments of the public. Instead, Americans perceive that their representatives pay substantial attention to their own supporters, economic elites, and campaign donors. This divergence between what the public prefers and what the public perceives about democratic representation explains public disapproval of congress, above and beyond the effects of other competing explanations such as gridlock, the quantity and quality of the laws passed, and personal and national financial well-being.

May 13, 2016
Concurrent Sessions E

The Future of Telephone Interviewing: Revealing New Data on the State of Telephone Surveys

Panel organizer: David Dutwin, *SSRS*
Trent D. Buskirk, *Marketing Systems Group*
Thomas M. Guterbock, *University of Virginia*

Telephonic survey research has been under considerable fire of late. Recent "misses" in election surveys at home and abroad have lead pollsters to question its efficacy. Costs have clearly increased, but there is little sense by how much and whether they will continue to rise. There have been virtually no research studies in the past ten years that have looked into whether the precipitous decline in response rates is associated with increased bias. And while there have been some studies that have aimed to document the relative quality of low response rate telephone samples to opt-in nonprobability internal panels, the data is still relatively sparse. This panel offers a wealth of new data designed to answer each

of these issues. Does telephone interviewing have a future? Our data shed considerable light on this critical question.

International Survey Measures, sponsored by WAPOR

Exploring a New Way to Detect Data Falsification in Surveys: Assessing the Meaning of "High Matches" between Observations in Survey Data

Katherine Simmons, *Pew Research Center*
Andrew Mercer, *Pew Research Center*
Steve Schwarzer, *Pew Research Center*

Concern about data falsification is as old as the profession of public opinion polling. However, the extent of data falsification, especially in international surveys, is not well documented. The standard approaches to address the problem include both prevention (e.g. extensive interviewer training) and detection (e.g. careful evaluation of patterns in data and paradata). A recent paper proposed a new detection method. The authors developed a tool that measures the maximum percentage of variables in a dataset on which an observation matches another observation. The novel claim in the paper is that valid survey data should have few observations that match on more than 85% of variables, what the authors term "near duplicates." We investigate how the measure of near duplicates performs alongside other traditional measures of data quality, such as the prevalence of extreme responses, item non-response, straight lining and section timings, to identify problematic interviews within a data set. We do so by expanding on the approach taken in Bredl, Winker and Kotschau (2008) of using multivariate techniques to identify interviewers with suspicious data patterns. Using datasets with known problematic cases, we evaluate how including a measure of near duplicates in this type of analysis affects how many, and which, interviewers are identified as needing further scrutiny. The paper discusses how different approaches to detecting falsification in datasets fit in the broader context of ensuring data quality in international surveys, and proposes some fruitful avenues for exploring how to improve standard practices in the future.

The Cross-national Equivalence of Customer Experience and Loyalty: The Case of Luxury Accommodation

Timothy B. Gravelle, *University of Essex*
Jill Andrews, *Four Seasons Hotels and Resorts*

In recent years, the survey methodology literature has paid increasing attention to the validity of cross-national comparisons. This has included increased focus on comparability in sampling procedures, translation of survey items to ensure comparability in meaning, and comparability in data collection procedures. Parallel to these developments, research on measurement equivalence (or measurement invariance) in cross-national survey research has been undertaken to evaluate statistically whether concepts are understood in the same way in different national, cultural, and linguistic contexts. This research has tended to focus on the measurement equivalence of concepts of interest to social and behavioral scientists – concepts such as attitudes toward democracy, national identity, the welfare state, attitudes toward immigration, or the structure of personality. To date, this literature has little engaged the concerns of researchers whose primary focus is customer survey data. This is a notable gap in the literature, especially considering that survey-based measures of customer experience and customer loyalty are important sources of information for decision-makers in large firms with global operations and global customer bases. As do many social scientists, managers with global firms typically

assume the cross-national equivalence of survey measures rather than test them. In the absence of such tests, it is possible that inaccurate conclusions about the performance of business units or customer perceptions may be drawn. This paper undertakes such a test, examining the measurement equivalence of customer experience and customer loyalty scores at a luxury hotel chain with a global presence. It uses multi-group confirmatory factor analysis (MGCFA) to determine whether such measures are invariant across languages and countries.

Age Heaping: Measuring Age Misreporting Across a Multinational Survey

Diana L. Greene, *RTI International*
Curtiss Cobb, *Facebook*
Annette M. Green, *RTI International*
Safaa Amer, *RTI International*

Age heaping is the tendency to round ages to the nearest 5 or 10 years. This measurement error affects the quality of age data and may affect whether individuals are classified as eligible to participate in a survey. While past literature has measured the level and correlates of age heaping in a single country, little is known about whether age heaping varies by country. Further, little is known about whether age heaping differs for self-reports versus proxy reports. To improve the understanding of age heaping across cultures and settings, we analyzed data from in-person surveys conducted in 12 lower and middle-income countries (n = 3,000 in each country). The surveys use national probability samples of people age 15-64 years old. We pursue two objectives. First, we examine the quality of age-related information collected through survey data by analyzing reported ages using the Whipple and Myers indices. These indices are well established methods for classifying the amount of age heaping. Using both indices we measure the levels of age heaping in each country. Second, using data collected in household and individual questionnaires, we examine variation across countries, focusing on proxy reporting versus self-reporting. We expect lower quality data from proxy reporting, as age is not always commonly known across family members. To investigate associations across countries we perform regression analysis on various respondent characteristics, including gender, education, literacy, and income. Preliminary results from Brazil, India, and Rwanda indicate that age heaping varies considerably across countries. The Whipple index suggest the ages reported were fairly accurate in Brazil but likely extremely inaccurate in India and Rwanda. In our paper, we consider the factors that contribute to age heaping and how these factors vary across lower and middle-income countries.

Exploring Feasibility of Mail Surveys in Urban China

Chan Zhang, *Fudan University*
James M. Lepkowski, *University of Michigan*

Mail surveys have been used in North America and Europe for decades and a large volume of research has been conducted to evaluate and improve mail survey design and data quality. However, mail surveys are still a new method yet to be explored in many other, often less developed countries. This study explores the feasibility of mail surveys in urban China. Mail surveys in China are expected to have better coverage than current landline only telephone surveys and are less intrusive than face-to-face interviews. Despite these potential advantages, mail surveys have been rarely used in China, mostly due to the common, but unexamined belief that few would respond. We conducted a 2 × 2 experiment in the fall of 2015 manipulating the delivery method (regular mail vs. courier delivery service) and the incentive payment (5 Yuan prepaid vs. no incentive). The sample included 2400 residential addresses systematically selected across 8 randomly selected neighborhoods in Shanghai. Courier delivery and

prepaid incentives had independent positive effects on response rates. Respondents across experimental conditions were more male and older than the general population, despite instructions included for random selection of the sample person in the household. The overrepresentation of males may be due to the within household selection instruction not being properly followed; the underrepresentation of younger respondents might reflect their preference for some more convenient response options. Overall, the study reveals the promise of mail surveys outside the Western context and points to unique challenges of conducting mail surveys in China.

The Impact of Survey Measurement Error: A Meta-analysis of Generalized Trust Research Using Either Dichtomous or Eleven-point Response Options
Sebastian Lundmark, *University of Gothenburg*

In survey methodology, it is frequently argued that increasing the number of response options in attitude questions, result in lower measurement error in subsequent operationalizations. In a recent POQ-article, generalized trust was found to—at the individual level—produce substantially lower measurement error when measured by an eleven-point scale compared to the otherwise standard dichotomous response scale. Despite this, measurement error at the individual level is in generalized trust research argued to not matter when using aggregates of survey responses for cross-national comparisons. This begs the question whether the impact of survey measurement error is perhaps exaggerated in survey methodology? The present study presents a meta-analysis of previous generalized trust research and compares the impact aggregates from the generalized trust question using either eleven-point or dichotomous response options have on cross-national level indicators. The paper finds that the eleven-point scale is strongly outperforming the dichotomy in explaining the country level outcomes. Hence, the higher measurement error in the dichotomous generalized trust question seems to carry over to aggregates and results at the national-level as well. And, although the dichotomy is sometimes able to create the expected relationship with the national level factors, the researcher runs a higher risk of substantially underestimating the impact generalized trust has on society, if s/he does not use the more correct measurement of the eleven-point scale.

Methodological Brief - Response Rates

Give It a Rest: Time between Call Attempts as a Method to Improve Telephone Survey Response Rates
Kelly Daley, *Abt SRBI*
Julie Pacer, *Abt SRBI*

Declining response rates in telephone surveys have prompted survey research organizations to design calling protocols that are cost efficient while balancing the need to achieve high survey response rates (Sangster 2012). These call protocols typically consider the best day(s) of the week to make calls, the maximum number of call attempts needed to obtain a completed interview and how much effort should go toward refusal conversion. However, regulatory and resource considerations for dialing cell phones can limit the number of call attempts made to cell phones. Furthermore, since most cell-phones have caller ID, frequent call attempts over a short time frame may over fatigue respondents, leading to a higher rate of refusals. With these considerations in mind, we designed a resting protocol on a national telephone survey. We hypothesized that a resting period in the middle of the calling protocol might improve overall response rates by increasing the likelihood of contact, where we could establish eligibility and also prevent hard refusals. This methodological brief reports the results of a resting

protocol on a national RDD cell phone survey of working adults, using call history data on a national dual frame RDD survey. The protocol included resting non-contact (busy, no answer, answering machine) sample for 14 days at the mid-point of the call design after the 4th call attempt (maximum attempts set to 8). We compare the final dispositions including contact and refusal rates between the sample that was rested (n=25,473) and sample that did not rest (n=18,598).

Optimizing the Effects of Online Survey Reminders

Tzu-Jou Wan, *American Institutes for Research*
HarmoniJoie Noel, *American Institutes for Research*
Lee Hargraves, *American Institutes for Research*

Reminders to online surveys are typically used to increase response rates. However, most of the literature focuses on a relatively short-term data collection period and a small number of reminders. This study looks at the effects of reminders sent during in a longer time period, i.e. 6 weeks from the initial invitation. Data were derived from the Qualified Health Plan Enrollee Experience Survey (QHP Enrollee Survey), funded by the Centers for Medicare and Medicaid Services. The QHP Enrollee Survey was an online survey administered in English to those who provided an email address when applying for health insurance through the Health Insurance Marketplace. There were 5 reminders sent through email after the initial invitation and each reminder was sent on every Monday throughout the field period. We answer the following research questions: 1) on average, how many reminders does it take to achieve a complete, 2) how long does the effect of a reminder last before there are diminishing completes, 3) are there differences in the demographic characteristics of early versus late respondents? It takes about 2-3 reminders on average to get a completed survey and, among QHP respondents, reminders are most effective within 2 days after the reminder. Male and those who aged between 35 and 54 are more likely to be encouraged by the 4th and/or the 5th reminder. This population might be considered nonrespondents if the field period was shortened. The findings from this study provide practical insights into the optimal number of reminders and time between reminders as well as how to target earlier versus later reminders to different types of respondents.

Is Shorter Better: Can a Follow-up Survey with Fewer Items Increase Response Without Impacting Data Quality?

Jocelyn Newsome, *Westat*
Jennifer Anderson, *Westat*
Kerry Yarkin Levin, *Westat*
Brenda Schafer, *Internal Revenue Service*
Patrick Langetieg, *Internal Revenue Service*

In the face of historically low response rates, researchers have explored whether shorter surveys can reduce nonresponse bias without compromising data quality. Although some studies have shown lengthier surveys encourage satisficing, item non-response, or increased "Don't Know" responses (Malhotra, 2008; Deuskens et al., 2004; Galesic & Bosnhjak, 2009), a 2011 meta-analysis conducted by Rolstad et al. found that the impact of questionnaire length on data quality was inconclusive. To test the impact of a shortened survey, we conducted an experiment with the IRS Individual Taxpayer Burden (ITB) Survey--an annual survey sent to 20,000 respondents that measures the time and money taxpayers spend complying with the tax law. The survey is comprised of two critical items that ask about time and money, along with 21 other items that provide context by asking generally about pre-filing and filing activities. The shortened version includes only the time and money items. An experiment conducted in

2013 suggested that the short form did not impact data quality. Surprisingly, the short form did not improve survey response. We hypothesized that response rates may not have improved because the short form was administered as the final contact, when it is more difficult to encourage response. In addition, both surveys were mailed in the same envelope, which meant that respondents might not have even realized they were receiving a shorter version of the questionnaire. In 2014, we redesigned the experiment. First, the short survey was sent earlier in the data collection process to non-respondents in population strata with low response (e.g., younger taxpayers with less complex returns). Second, to distinguish the shortened survey from other mailings, it was sent in a smaller envelope. Results will examine the short version's impact on overall response rates, as well as discuss its impact on data quality.

How to Invite? Methods for Increasing Internet Survey Response Rate

Alison R. Huang, *American Institutes for Research*
HarmoniJoie Noel, *American Institutes for Research*
Lee Hargraves, *American Institutes for Research*

While internet surveys garner historically lower response rates compared to other survey modes, strategic design of initial respondent contact can help boost response. Mail pre-notification letters may aid in legitimizing a mailed or internet questionnaire, however there are additional costs associated with mailed pre-notification letters. At the same time, there are concerns about e-mailed pre-notification letters being viewed as spam, yet one may expect that an invitation in the same mode as the survey may encourage response. This methodological brief will show the results of a randomized pre-notification letter mode experiment embedded in a web survey to reveal differences in response rates and respondent characteristics between mail and e-mail pre-notification conditions. The Qualified Health Plan Enrollee Experience Survey, funded by the Centers for Medicaid and Medicare Services, gathers information on people's experiences with the health plan they enrolled in through the Affordable Care Act's Marketplaces. During the survey field test in 2014, 3,340 people were randomly assigned into two equal groups. One group received the pre-notification invitation by e-mail while the other received the invitation by mail with instructions to complete the survey online. All subsequent reminders were sent via e-mail. Analyses showed that the response rate was 1.4 times higher for those who received a mailed pre-notification invitation (16.0%) compared to an e-mailed invitation (11.3%). The results of this mode experiment are important to informing data collection methods for researchers considering an internet option as part of their data collection design.

Can a Locally-tailored Image Improve Return Rates? Evidence from a Community-specific Study in California

Matt Jans, *UCLA Center for Health Policy Research*
Julie C. Linville, *RTI, International*
David Grant, *University of California, Los Angeles*

Envelope graphics tend to reduce response rates in mail surveys, but locally-iconic or neighborhood-relevant images may improve them by activating local affiliation and identification. We tested this hypothesis experimentally in 14 California communities (about 28,500 Delivery Sequence File addresses; about 2,000 per community). The response tasks were returning a household information form with a phone number, a screener interview by phone, and an extended phone interview. Within each community, addresses were randomly assigned to receive either a standard envelope with no image (other than the sponsor logo in the return address field), or an envelope that also had a locally-

tailored image. Using Google image search, we chose images that we thought would be locally-iconic (e.g., notable building or landmark, train station, local landscape, or city sign). We used a community logo from a local health improvement program when community liaisons requested or we found no better image. Local photographic images were used in 9 communities and the logo was used in 5 communities. Initial results show a statistically-significant (alpha < 0.05) negative effect of about 1 percentage point on household information form return rate when using the image/logo. The effect seems due to one community with a 3.5 percentage point difference. There was no overall effect of the image/logo on phone screener completion or extended phone interview completion, but two communities had marginally-significantly (alpha < 0.1) lower phone screener completion rates with the image, each about 1 percentage point. Three had marginally-lower and one had a significantly-lower extended interview completion rate of about 1 percentage point with the image. Our findings support traditional advice that adorning mailing materials with images and logos can have negative effects on completion rates and should be avoided unless their effects can be tested beforehand.

Nonresponse Bias for Survey Estimates of Social Activities and Roles
Ashley Amaya, *University of Maryland, JPSM*

Nonresponse bias is a fundamental concern for survey researchers as understanding when and to what extent it occurs is critical to producing accurate statistics. According to the social integration hypothesis, individuals who participate in a broad range of social activities and roles should be more likely to respond to surveys (Goyder 1987; Groves & Couper 1998). As a result, prevalence estimates of social activities and roles should be upwardly biased. By contrast, models predicting these activities and roles may be unbiased if the nonrespondents are missing at random, as the results of Abraham, Helms, & Presser 2009 suggest. Using the rich frame information available on the American Time Use Survey (ATUS) and the Survey of Health, Ageing, and Retirement in Europe (SHARE) Wave II, we compare the full sample to the respondent sample on 30 different social roles and activities. The results suggest that nonresponse bias was widespread and often large on univariate estimates, but was frequently small and typically did not alter the inferences drawn from multivariate models.

What Are You? Measuring the Size, Characteristics and Attitudes of the Multiracial Population in America

Panel organizer: Richard Morin, *Pew Research Cener*

The multiracial population in the United States is growing three times as fast as the population as a whole. These Americans represent the cutting edge of a demographic trend that is poised to transform the way the country thinks about race and how individuals define their own racial identity.Yet this population remains under-studied, in large part because of the major practical and conceptual challenges that researchers confront when they attempt to measure the mixed-race population. This panel explores a number of those issues, ranging from conflicting definitions of who is multiracial to the challenges posed by shifting concepts of race and ethnic identity. In addition, panelists will present data on the size of the multiracial population, the percentage of adults with a mixed-race background who do not consider themselves to be multiracial, the share of Latinos who consider "Hispanic" to be a race and also summarize the latest findings of ongoing experiments by the U.S. Census Bureau that attempt to better measure race and ethnicity.

Exploring New Ways to Measure Race/Ethnicity for the 2020 Census
Nicholas Jones, *U.S. Census Bureau*

Nicholas A. Jones, Director of Race and Ethnic Research and Outreach in the U.S. Census Bureau's Population Division, summarizes his team's research to improve the measurement of race/ethnic data for the 2020 Census. He will present an update on the 2015 National Content Test (NCT) research on race and ethnicity, which is the Census Bureau's primary mid-decade study for developing recommendations for the 2020 Census' race/ethnicity question(s). He will explain how this survey of 1.2 million households and its corresponding reinterview provide the basis for making critical research recommendations. He will describe the key dimensions of the NCT research, which focus on testing alternative question formats (separate question approach vs. combined question approach); testing a potential "Middle Eastern or North African" category; testing alternative instructions and terminology to improve reporting, especially for multiracial/multi-ethnic respondents; and utilizing new question designs to elicit the reporting of detailed groups. The research effort aims to help provide the most accurate, reliable, and relevant race/e-thnicity data possible about our changing and diversifying nation.

Measuring Hispanic Racial Identity: A Challenge to Traditional Definitions of Race
Mark Hugo Lopez, *Pew Research Center*
Ana Gonzalez-Barrera, *Pew Research Center*

Hispanics view their racial identity differently than most other Americans and this presentation will highlight some of the major differences. Based on a Pew Research Center survey of 2,438 Hispanics in 2015, about two-thirds of all Latinos consider "Hispanic" to be their race or their race and ethnicity. This attitude stands in sharp contrast to the traditional federal government definition of "Hispanic," which considers Hispanic ancestry to be an ethnic identity and not a race. The disconnect between the government's definition and how Hispanics view themselves has complicated federal efforts to accurately measure the Hispanic population. This presentation reports more fully on the results of the 2015 study and on the findings of the 2015 and 2014 National Survey of Latinos (NSL). The 2015 NSL asked half of its nationally representative sample of 1,500 Hispanics to report their race in the traditional way (not including a Hispanic or Latino category but allowing for volunteered responses), the other half of the sample was asked to report their race using a question similar to the AQE experiment by U.S. Census Bureau, which includes the Hispanic category alongside other races. The 2014 NSL also explored whether Hispanics had background with roots among the indigenous peoples of the Americas; it explored mixed-race identity among Latinos using mixed-race terms such as mestizo or mulatto, terms that are used in countries like Mexico to capture a mixed indigenous and European heritage.

The Size, Characteristics and Key Attitudes of Multiracial Americans
Juliana Horowitz, *Pew Research Center*
Richard Morin, *Pew Research Center*

This presentation will summarize the findings of a major national survey of multiracial Americans by the Pew Research Center in 2015, including some of the methodological issues encountered during the project. Using a different method than the U.S. Census for measuring race, the Pew Research survey found that 6.9% of American adults have a mixed race background, or about three times larger the U.S. Census estimate. To determine racial background, researchers asked respondents to identify their race or races and then asked the race/s of their parents and grandparents. This

technique produced a larger share of multiracial Americans but also captured a significant number of individuals with a mixed race background who, when asked later in the survey, said they did not identify as multiracial. In fact, about six-in ten adults with mixed-race ancestry said they did not consider themselves to be multiracial. The major reasons why they did not: They were raised as a single race, they identified as a single race or they looked like a single race. Together, these findings suggest that multiracial identity is as much an attitude as a biological or genetic fact. Americans with a background that includes white and Native American ancestors make up about half of all mixed race adults. In terms of their attitudes, multiracial adults are proud of their mixed-race background (60%) and feel their racial heritage has made them more open to other cultures (59%). At the same time, a majority (55%) say they have been subjected to racial slurs or jokes, and about one-in-four (24%) have felt annoyed because people have made assumptions about their race. Still, few see their multiracial background as a liability. In fact, only 4% say having a mixed racial background has been a disadvantage in their life.

Do Multiracial Adults Favor One of their Background Races Over the Other: An Implicit Association Test
Annie Franco, *Stanford University*

To test how biracial adults view their background races, the Pew Research Center and a team of researchers led by Stanford University professor Shanto Iyengar conducted an Implicit Association Test (IAT) with representative samples of biracial white-black and white-Asian adults as well as with samples of single-race black, white and Asian adults. The samples were drawn from a YouGov panel. The IAT measures subconscious or "hidden" bias by how quickly test subjects associate good and bad words with specific racial groups. The test found that about three-quarters of all adults demonstrated various degrees of racial bias, including biracial adults. About four-in-ten biracial black-white adults showed an implicit preference for whites over blacks while about a third preferred blacks over whites. This stands in contrast to their responses to explicit measures of group affect, on which biracial Blacks tend to align with blacks rather than whites. Biracial white-Asian adults were more evenly divided in their implicit preferences, a finding consistent with their responses to the explicit measures: 39% favored whites over Asians while 38% preferred Asians. This presentation explores the implications of these results and how the IAT can complement more traditional forms of survey research.

50 years of the National Longitudinal Surveys Program: Reflections on the Last Half-Century of Large-scale High-quality Survey Research

Panel organizer: A. Rupa Datta, *NORC at the University of Chicago*

The National Longitudinal Surveys Program of the Bureau of Labor Statistics, U.S. Department of Labor, is celebrating its fiftieth anniversary. Tracing survey operations and methods over the life of the program also provides a lens into the evolution of survey research over this period. The NLS program has included more than 50,000 individuals across seven cohorts born throughout the 20th century. These individuals have participated in more than 100 survey rounds, producing almost 700,000 interviews and assessments documenting individual's employment-related experiences in such diverse areas as education, physical and mental health, family life, labor force experience, cognition, income and assets and migration. Data for the earliest cohorts were collected by the Census Bureau, while NORC at the University of Chicago has collected data for cohorts initiated in 1979 and 1997. CHRR at The Ohio State University has been involved in all NLS cohorts. This session discusses several aspects of the NLS program over the course of its history, reflecting on changes in the social and policy context,

operational systems and processes, and distinctive challenges and accomplishments. These aspects include: 1) sampling and weighting approaches in nationally representative area probability surveys, 2) conducting and managing data collection, 3) training field interviewers, 4) questionnaire design in longitudinal surveys, 5) data quality assurance, and 6) data delivery and dissemination. The NLS cohorts began with paper-and-pencil in-person interviews, but were early adopters of Computer-assisted Personal Interviewing, Audio Computer-Assisted Self Interviewing, and distributed computer-assisted telephone interviewing. Behind the scenes, survey managers can now see and analyze real-time case-level call logs, while managers of 40 years ago received only weekly counts of each interviewer's completes and non-interviews. The session promises to be a thoughtful tour from decks and cards through 5" floppies to web dissemination and social media!

Adapting Weighting Procedures to Complex Sample Design in NLSY97
Colm O'Muircheartaigh, *University of Chicago*
Steven Pedlow, *NORC at the University of Chicago*

Each cohort of the NLS program has included some oversamples, although the types of populations oversampled and the methods of oversampling have varied over time. This paper presents a method of weighting for NLSY97 that substantially increased the precision of the estimates over that obtained by the standard method. The sample design for NLSY97 incorporated a supplemental oversample of individuals from areas with unusually high density of African-American and Hispanic populations. The supplemental sample was selected independently of the main (national cross-section) sample. Decisions on how to adapt estimates from these two independent samples to the complex sample design had important implications for the precision of estimates. To combine data from the main and supplemental samples, the standard method is to carry out estimation separately for the two samples, then combine the estimates to get a composite estimate based on all the data. The variance of the composite estimate is obtained by adding the variances of the separate estimates, each weighted by the inverse of its precision, where the weights add to 1. This is the precision-weighted variance, and treats each of the two samples as a unit. This may be called the combining samples approach. The alternative method developed exploits the idea of first combining the sample into a single sample by calculating the overall probability of selection of each individual i.e. the probability of an individual's being selected into either of the two samples. By treating the data as a single sample, the variance can be calculated directly. This is the combining cases approach. The theory for the two approaches is presented, together with the empirical results. The combining cases approach demonstrated quite substantial increases in precision, ranging from 10% for the whole sample to 40% for some minority estimates.

Designing Longitudinal Labor Force Survey Questionnaires 1965 to the Present
A. Rupa Datta, *NORC at the University of Chicago*

Throughout the 50 year life of the NLS program, the purpose of the surveys has been constant: to better understand the determinants and consequences of labor market experience. The actual questions asked have varied by the life stage of the respondent (ranging from ages 12 to 84), the policy and economic environment at the time, the pressing research questions anticipated for the future, and the technical capacities of the survey process. In this respect, the story of the NLSs is generally a story of dramatically increasing data quality and vastly expanding frontiers of the research questions that can be asked and answered with the data. The earliest NLS questionnaires are essentially repeated point-in-time data captures administered to a longitudinal sample. Activities between interview dates are rarely documented except for such key events as job starts and stops or a spouse's death. One innovation in

the NLS design was to seek to interview individuals who had not participated in the prior round; at the time, the convention was to only interview remaining panel members. A second innovation was to collect more information in an event history format, "Since the date of last interview, have you…." These two adaptations together increased the longitudinal nature of the data collected and diminished loss of data due to sample attrition, since someone missing an interview could simply report major milestones in the subsequent interview. Computer-assisted interviewing expanded the types of data collected and the quality of the data collected, and therefore the range of research that could be conducted. Specific methods for improvement and associated research opportunities are reviewed.

Evolution of Survey Operations in the NLSYs
Lauren Seward, *NORC at the University of Chicago*

While the National Longitudinal Surveys have looked at the change or lack of change in employment and other areas within the lives of its respondents over time, it is time to turn the focus inward. As the NLS celebrates its 50th anniversary this year, it is time to take a look back at how the history of data collection, sample and field management and our interactions with our respondents have changed over the long history of the study and what impact it has had on this and other studies. What changes have occurred in how its data has been collected and how the study has been fielded? In what ways has innovations in technologies and new ways of collecting data allowed the fielding of the NLS to evolve over its history? This paper will focus on the history and progression of data collection activities of the active cohorts of the NLS surveys (NLSY79 and NLSY97) including recruiting and field training, field and sample management, data collection, and the evolution of the different channels in maintaining connections, interactions and communications with our respondent and how its history is paving the way for future innovations and enhancements to the areas to help sustain an almost 80% response rate of living respondents in both cohorts.

Incorporating Technology and Adult Learning Methods into Field Interviewer Trainings
Vicki Wilmer, *NORC at the University of Chicago*

The methods for training field interviewers and field managers has evolved quite a bit over the years. As expensive, multi-day, in person trainings have given way to online learning management systems, the NORC team has implemented innovative efficiencies in order to provide quality training to the field data collectors. As part of the NLS at 50 panel, we will describe the training that was conducted on the earlier rounds of the NLSY79 and 97, including in person overview of the questionnaires, the case management system, and certification of all instruments, especially the Child PIAT and PPVT assessments. As technology advances and budgets tighten, we continue to look for ways to offer a quality training at a lower cost. In the past two decades, this includes moving towards remote training and certification. This has included conference calls, narrated videos, web tutorials, booster-training modules, use of field memos, online learning management systems, and "live" sessions in which trainers remotely connect to trainee machines in order to observe progress. Yet, despite a physical distance between trainers and trainees, we find that data quality has remained largely intact, with the number of falsified cases now almost non-existent, and the percentage of high data quality scores steady. Using a foundation of quality, in person training, it is possible to deliver the same components in a myriad of remote methods that utilize technology and cost efficiency.

Recordings, Reinterviews and Review: Data Quality Assurance Across the Years
Isabella Velasquez, *NORC at the Unversity of Chicago*

Excellent data quality is critical to longitudinal research. Efforts to monitor data quality in the NLS program have varied with best practices over time and given available resources. Historical data quality assurance activities have included supervisors accompanying interviewers on selected interview visits, telephone reinterviews with selected respondents, audiotaping of interviews that were then reviewed by survey management staff, hand-editing of completed paper questionnaires by field supervisors, and occasional retrieval of missing information through follow-up contacts. In paper-and-pencil days, additional quality assurance ensured that interviewers were accurately following skip instructions, and that transcription errors were not introduced during data entry. Today's data quality assurance processes in the NLS program reflect our times. Processes are distinguished from earlier efforts by being more formally integrated into an on-going system of training and continuous improvement, and by employing a combination of statistical approaches to identify potential anomalies for investigation. Data quality assurance begins with a computerized interview with many built in checks and balances, as well as our rigorous training, which emphasizes the fundamentals of data quality. Since 2008, the project has used computer-assisted recorded interviewing as its primary mechanism for verifying interviewers' adherence to protocol by grading interview recordings based on certain criteria. Feedback on randomly selected interviews inspected at randomly selected items are then shared with interviewers. If needed, interviewers participate in remedial coaching sessions. We supplement recording review through statistical data review. Statistical comparisons across rounds within respondents, across responses within rounds, and across interviewer caseloads identify potentially suspect cases or interviewers for closer manual inspection and review. A final yet vital test is the "gut feeling" of our Field Managers. By combining the intuition of our field staff with modern monitoring tools, we are able to retain the high quality standards of the NLS.

Beware of What You Wish For: Navigating the Challenges of 50 Years of Data Dissemination
Rosella Gardecki, *The Ohio State University*

Data dissemination may be the aspect of survey operations that has been most directly upended by five decades' developments in information technology. The fundamental principles that governed early NLS data files are essentially universally inapplicable today: 1) questionnaires were simple, easy to follow, and the best documentation of a data file, 2) only very few high-skill/high-sophistication researchers could access the data which limited research but had the unintended byproduct of providing a certain level of data security, and 3) data storage was expensive. This paper discusses three components of data dissemination (documentation, the trade-between data availability and data security, file size) in light of dramatic changes that have taken place as well as the new challenges that have arisen in their stead. While computer-assisted interviewing has vastly expanded our ability to flexibly and efficiently collect data in respondent-friendly ways, more complex interviews demand tools that can help data users understand and navigate data files as well as questionnaires once did. (Indeed, the first NLS questionnaire was approximately 20 pages, while a current questionnaire would require more than 1,000 pages if printed.) Web availability of public use data files along with statistical functions in ordinary home computing software have dramatically increased effective access to data files. Together with widespread publication of other online resources, the potential for deductive disclosure has increased exponentially, in ways that can be difficult to anticipate and protect against. Finally, advances in statistical computing have mostly been boons for the NLS: longer variables names, more flexible variable labels, freedom from 'decks and cards', essentially unlimited data file size. At the same

time, colossal numbers of variables generate the need for new mechanisms of variable search and variable exploration and open up the potential for researcher confusion as never before.

Preferences and Outcomes in Multi-Mode Research

Effects of an Initial Offering of Multiple Survey Response Options on Response Rates

Elizabeth A. Steele, *The Gallup Organization*
Jenny Marlar, *The Gallup Organization*
Laura Allen, *The Gallup Organization*
Kirti N. Kanitkar, *The Gallup Organization*

Initial research into mixed-mode methods suggested that characteristics of nonrespondents might vary by data collection mode (Groves 2006). However, much of the research done on mixed mode studies has involved using multiple modes during multiple stages of contact, confounding mixed-modes methods with multiple attempt strategies (Beebe et al. 2012). There is limited recent research on offering respondents a variety of modes for survey completion at the initial contact in nationally representative studies. Research completed in 2015 suggests that respondents are more likely to respond via web relative to mail and fax (Thompson, Oliver and Beck, 2015.) Given the explosion of web access in the last five years, this study will evaluate current preferences for completion mode. This study will explore the effect of offering multiple response modes during initial contact with respondents. Analysis will be done on the response rate associated with the number of modes offered and the type of modes offered. Special attention will be paid to the demo graphic groups that respond in each mode. In 2015, Gallup fielded a survey of the general population using a mail mode as the initial contact type. Within this initial mail contact there are 5 possible treatments: (1) a mailed paper survey (2) a mailed paper survey with a web link (3) a mailed web link with an option to request a paper survey (4) a mailed paper survey with a QR and (5) a mailed paper survey with both a web link and a QR code.

Mode Effects on Subjective Well-being Research: Do they Affect Regression Coefficients?

Rosa Sanchez Tome, *University of Lausanne*
Caroline Roberts, *University of Lausanne*
Michèle Ernst Stähli, *University of Lausanne*
Dominique Joye, *University of Lausanne*

Indicators of subjective well-being have become essential to measure prosperity and societal progress across the world, with much of the produced research being based on survey data at a time in which traditional survey modes have trouble maintaining data quality and costs. One of the most widespread alternatives in overcoming data collection challenges is to use various modes of data collection but as a consequence, there is great concern that data collected through different modes may not be comparable. Using data from a mode comparison experiment implemented in Switzerland, we compared telephone, mail and web modes and found both selection and measurement effects for some widely used variables such as happiness and job satisfaction. In this paper, we examine the impact that mode effects have on the statistical relationship between variables. To do this, we replicate OLS multiple regression analyses that are widely used by researchers to predict happiness, satisfaction at work and social trust for each mode of data collection, with and without controlling for selection effects. The regressions are implemented for three different scenarios depending on whether mode effects were

found a) on the dependent variable but not on independent variables; b) on independent variables, but not on the dependent variable; and c) on both dependent and independent variables. Results show differences in regression coefficients and the relationship between the variables, indicating that mode effects play an important role on the conclusions obtained in cross-sectional research into well-being.

Mobile Device Use in Web Surveys Among College Students: Predictors and Consequences for Data Quality

Scott Beach, *University of Pittsburgh*
Donald Musa, *University of Pittsburgh*
Stephen Strotmeyer, *University of Pittsburgh*
Janet Schlarb, *University of Pittsburgh*

Increasing numbers of surveys are completed on mobile devices like Smartphones, and questions have been raised about the quality of mobile survey data versus data collected by laptop / PC. This paper examines predictors of mobile device use and the consequences of this choice for data quality in web surveys among undergraduates at the University of Pittsburgh (U.S.). The data come from an ongoing panel survey that has involved changes in the survey protocol over time and experimental manipulations related to device choice to explore data quality. This analysis incorporates combined data from 2013 (n=1,626), 2014 (n=1,431), 2015 (n=3,077), and 2016 (anticipated n~3,100; total n~ 9,200). Multivariate models with robust standard errors (for repeated individual observations) are tested for the following outcomes: choice of mobile vs. non-mobile device; and the following data quality indicators: (1) survey break-offs; (2) item missing data; (3) survey completion time; (4) provision of open-ended text responses; (5) length of text responses; and (6) substantive survey outcomes (i.e., "mode effects"). Predictors include survey year, class (freshman, sophomore, junior), sex, race, SAT score, school, and survey history (# prior invitations / completions, prior devices used). The key predictor for the data quality models is use of mobile vs. non-mobile device. Two survey design factors that have varied over time will also be included as predictors: (1) survey format - mobile (no grids) vs. non-mobile (grids); (2) explicit mention in email invitation that the survey can be taken on mobile or non-mobile devices (vs. no mention). Main effects and interaction effects of predictor variables by type of device used will be tested. The goal of the paper is to use a large data set with multiple covariates to provide insights into the predictors and effects of mobile device use in web surveys.

Mode Effects in American Trends Panel: Bayesian Analysis of a Cross-classified Item-person Mixed Model

Jeff Gill, *Washington University at St. Louis*
Stanislav Kolenikov, *Abt SRBI*
Kyley McGeeney, *Pew Research Center`*

American Trends Panel is a probability panel with RDD recruitment developed by Pew Research Center and Abt SRBI. Over the life of the panel, surveys have been conducted primarily via web mode, with mail mode for those who do not have access to the Internet or do not provide an email address. We analyze the results of the July 2014 wave (Wave 5) that included a comprehensive, large-scale mode-of-interview experiment that randomly assigned web respondents to telephone and web modes, with approximately 1,500 respondents in each mode. To quantify the contributions to the mode effects of the different question characteristics in the 75-question instrument, we build a cross-classified model with effects of person and question characteristics to identify the properties of survey questions that make them susceptible to mode effects, as well as the demographic groups that tend to exhibit mode

effects. The model was estimated by Markov chain Monte Carlo computational Bayesian methods using a combination of R and JAGS packages. We discuss how the magnitude of the mode effect is associated with the properties of the survey questions, such as question format, topic, the potential of social desirability; and with personal characteristics, such as education, race, and the level of political engagement.

A Test of Web/PAPI Protocols and Incentives for the Residential Energy Consumption Survey

Paul P. Biemer, *RTI International*
Joe Murphy, *RTI International*
Stephanie Zimmer, *RTI International*
James (Chip) Berry, *U.S. Energy Information Administration*
Katie Lewis, *U.S. Energy Information Administration*
Deng Shaofen, *U.S. Energy Information Administration*

The Residential Energy Consumption Survey (RECS), sponsored by EIA, collects data on householder behaviors and housing characteristics that affect current and long-term energy usage and cost. To build a more responsive and cost-effective data platform, EIA is considering moving the RECS from the current personal interview mode to a Web/PAPI mixed-mode survey design. This paper describes an experiment to test two incentive options and four data collection protocols arrayed in 2x4 factorial design. A national epsem sample of 9,650 households was selected and divided equally among the eight treatments. The two incentive treatments included a $5 prepaid incentive in the first questionnaire mailing but varied the promised incentive amount from $10 to $20 paid after the completed questionnaire is received. The four protocol treatments are: (a) CAWI (or Web only), (b) Choice (which gives respondents the simultaneous choice of responding by either PAPI or Web), (c) Choice+ (similar to the Choice protocol but adds a $10 bonus incentive if the respondent chooses to respond by Web) and (d) CAWI/PAPI (which initially offers the Web only option, but allows the choice of Web or PAPI in nonresponse follow-up invitations). The paper reports the results of the study including which data collection protocol/incentive structure combination provides the highest quality estimates based upon a total survey error analysis.

Mode Effect on Racial Sensitive Questions between Web and Computer-assisted Self-interview

Mingnan Liu, *SurveyMonkey*
Yichen Wang, *NERA Economic Consulting*
James M. Lepkowski, *University of Michigan*

As surveys are collecting more and more sensitive information, researchers will adopt more self-administered data collection mode or module. Previous study has compared the mode effect on sensitive questions between self-administered and interviewer-administered modes. In general, they concluded that with self-administration, respondents were more likely to disclose sensitive information and gave undesirable responses. However, to our knowledge, very limited effort has been devoted to comparing different self-administered modes on sensitive questions. The 2012 American National Election Study was conducted in two self-administered modes, one by Web and another through computer-assisted self-interview (CASI). We examine the mode effect on a set of race-relevant questions, which are believed to be sensitive and can trigger socially desirable answers. The preliminary results show a mixed finding of mode effect. While Web elicits more sensitive answers to some

questions, CASI receive more such answers to other questions. In addition, we find different patterns of mode effect from different racial and ethnic groups. In the presentation, we will summarize the previous study, provide our hypotheses, present the final results on mode effect, the interaction between mode and respondent's race/ethnicity, and discuss the implication on future study and survey practice.

Writing Questions for Web Surveys: Cutting Edge Issues in Design & Implementation

Conducting Survey Experiments Using an Online Labor Market

Stephanie Fowler, *National Cancer Institute*
Gordon Willis, *National Cancer Institute*
Richard P. Moser, *National Cancer Institute*
Reanne L.M. Townsend, *Westat*
Aaron Maitland, *Westat*
Hanyu Sun, *Westat*
Rebecca Ferrer, *National Cancer Institute*
David Berrigan, *National Cancer Institute*

Questionnaire design, pretesting, and evaluation can be enhanced by the use of planned experiments that manipulate variables like question phrasing and ordering, and that assess test-retest reliability. However, experimentation involving survey questions has been limited because of cost, burden, and time constraints in obtaining appropriate samples. This presentation will describe the use of Amazon's Mechanical Turk (MTurk), an online labor market, to examine question order effects, test-retest reliability, and the potential utility of an online crowdsourcing platform for the conduct of survey experiments. We selected survey self-report items on perceptions of the environment related to walking, and on the frequency and duration of walking for transportation and leisure, from the 2015 Cancer Control Supplement of the National Health Interview Survey (NHIS). A total of 1,446 respondents were recruited through MTurk and completed the web survey using Qualtrics. At Time 1, respondents were randomly assigned to one of two conditions, in which the order of questions on walking behavior, versus perceptions of the walking environment, was varied. Four weeks later, respondents were invited to complete the survey in the same order as in Time 1. We obtained a wide range of demographic characteristics (56.2% male, 84% Non-White Hispanic, 50% with at least a Bachelor's degree, and mean age of 31.8 (SD = 10.6). Results demonstrated reasonable item test-retest reliability for the items administered (Kappa = .48-.84). Further, as in past studies of walking-related survey items, we observed a significant question order effect. The total cost for respondent incentives and MTurk administrative fees was $750. Based on success in carrying out the operational aspects of the experiment in a quick, low-cost manner, and in obtaining information useful for driving questionnaire-design decisions, we conclude that an online platform like MTurk is useful for embedding survey experiments.

Cheating on Political Knowledge Questions in Online Surveys: An Assessment of the Problem and Solutions

Scott Clifford, *University of Houston*
Jennifer Jerit, *Stony Brook University*

The use of online surveys has increased dramatically, but concerns remain about the quality of data produced by respondents from a remote location. One such concern is that respondents might use the web to look up answers to performance questions, such as political knowledge, invalidating their

responses. Existing evidence for the prevalence of "cheating" behavior is unclear and some scholars are skeptical that respondents are sufficiently motivated to cheat in an anonymous online survey. In contrast, we argue that respondents may be motivated by self-deceptive enhancement, and searching the web provides an opportunity to provide the correct answer while feeling as if they knew it all along. In this paper, we first provide evidence that such cheating behavior is widespread, but its prevalence varies considerably by sample. Next, we experimentally test a variety of methods for reducing cheating. Our results demonstrate that direct requests to not cheat are ineffective. Time limits proved more effective, but there is some evidence they interfere with the usual answer process. Our evidence suggests that a commitment mechanism, which involves asking respondents to agree to not cheat, proved the most effective. Finally, we demonstrate that the commitment treatment reduces the positive correlation between political knowledge scores and self-deceptive enhancement, providing evidence for the motivation behind cheating behavior. This finding also demonstrates that the commitment mechanism improves the discriminant validity of political knowledge measures. Overall, our paper clarifies the debate over cheating behavior in web-based surveys and provides researchers with a method for coping with the problem.

Assessing the Effects and Effectiveness of Attention-check Questions in Web Surveys: Evidence From a 14 Country Cross-national Survey Experiment
David L. Vannette, *Qualtrics*

The quality of survey responses in web surveys often leaves much to be desired. Respondents commonly engage in negative response behaviors and take shortcuts in the ideal response process that researchers would like them to follow when answering questions. The use of attention-check questions in web surveys has become one extremely popular method that has attempted to identify some of these inattentive or misbehaving respondents and exclude them from either analysis or further participation in the survey. In this study, we report the results of experimental interventions aimed at increasing respondent attention, commitment and thoughtfulness. Unlike the production survey context, the respondents that failed the attention-check or commitment questions were allowed to remain in the sample to evaluate how they influence the data. The effects of these interventions on data quality are assessed using outcome measures that include the quantity and quality of text generated in open-ended responses, response times to questions, socially desirable responding, and non-differentiation in responses. The data for this study were collected, using the Qualtrics web survey platform, from diverse populations of web survey respondents in 14 independently sampled countries. The results from this study will inform knowledge about the effectiveness of interventions aimed at improving the quality of responses in web surveys and will be of value to all researchers using this method.

Investigating Cognitive Effort of Response Formats in Web Surveys using Paradata
Jan Karem Höhne, *University of Göttingen*
Stephan Schlosser, *University of Göttingen*
Dagmar Krebs, *University of Giessen*

Since Likert's (1932) eminent article "A Technique for the Measurement of Attitudes", the use of agree/disagree questions has become increasingly popular in empirical social research, because it appears to be possible to measure several constructs with a single response format. Fowler (1995), by contrast, argues that using construct-specific questions represents a simpler, more direct, and more informative method than using agree/disagree questions. Until now, this assumption has never been

tested empirically. In this study, we investigate the cognitive effort of agree/disagree and construct-specific questions using paradata. By measuring response times, we are able to investigate the cognitive information processing of respondents during web surveys to determine the burden of these response formats. We applied an innovative two-stage outlier correction that is based on the activity of the web survey while processing, accompanied by an outlier definition that is based on the response time distributions. Moreover, computer mouse clicks were captured to evaluate the response times. We conducted an experimental study with four groups (n1 = 255; n2 = 237; n3 = 278; n4 = 235) that is based on an onomastic sampling approach. We tested agree/disagree and construct-specific questions, both presented individually as well as in grids. The results of our investigations indicate that the single construct-specific questions produce significantly higher response times than their agree/disagree counterparts. Hence, it is to be assumed that they imply a higher cognitive effort. However, there are no differences between the grids. In addition, the mouse clicks show no significant differences between the groups; thus, it is to be expected that they do not affect response times. Altogether, it appears that construct-specific questions require deeper cognitive information processing than agree/disagree questions if they are presented individually.

Using Paradata to Identify Questions with High Respondent Burden for Improvement in Future Surveys

Rebecca J. Powell, *RTI International*
Ashley Richards, *RTI International*
Shengchao Yu, *New York City Department of Health and Mental Hygiene*
Robert Brackbill, *New York City Department of Health and Mental Hygiene*

Web surveys are becoming increasingly popular among researchers due to many benefits such as reduced costs, efficiency of data entry, and automated navigation of the survey for the respondent. Researchers need to consider the design of the questions and web page to reduce respondent burden, which has been shown to increase breakoff rates and decrease response quality overall. This study aims to understand how respondent burden varies across different types of questions by using paradata from the World Trade Center Health Registry's (WTCHR) Wave 4 survey. In particular, we compare breakoff rates, time latencies, and number of navigational backups to identify questions that are particularly burdensome for some respondents. Preliminary findings indicate that 20% of breakoffs occurred on pages displaying more than one question, even though the vast majority of the survey pages displayed only one question. The majority of those screens with multiple questions contained grid questions. We further examine the relationships between these paradata measures and other characteristics of the survey, such as question content, question type, and survey length. Additionally, we assess differences in these paradata measures across demographic groups and device used to take the survey to help differentiate between issues due to the question and respondent characteristics. The results from this study will be used to (1) reduce respondent burden, decrease item nonresponse and therefore improve accuracy of the data for future WTCHR surveys, and (2) inform other survey researchers of ways to reduce respondent burden and improve response quality in their own surveys.

Public Opinion on Income Inequality and the Economy

Lagging Indicators of the Great Recession

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

Official business cycle measures indicate that the Great Recession ended in 2009. But numerous public assessments of both the general economy and their personal economic and socio-economic position, indicated little or no improvement through at least 2012. Some indicators even showed worsening evlauations through 2012 and no improvement in 2014-15. This paper looks at trends in assessments of general and personal economic and societal well-being before, during, and following the 2007-09 Great Recession. Besides using official governmental indicators of GNP/GDP, unemployment, laborforce participation, and similar measures, this paper looks at public assessemnt from both cross-sectional and panel surveys. These include NORC's General Social Surveys, Pew surveys, Gallup surveys, and others. The paper evaluates what indicators improved more or less rapidly and why some indicators were lagging indicators.

When Do the Rich Win?

James Alexander Branham, *University of Texas at Austin*
Stuart Soroka, *University of Michigan*
Christopher Wlezien, *University of Texas at Austin*

There is growing academic and non-academic interest in the relative influence of economic haves and have-nots on public policy. In a recent, well-known article, Gilens and Page (2014) compare the influence of upper and middle-income citizens and find that the preferences of the former are all that matter for policy representation. Here, we build on this work, exploring just how often the rich win and the kinds of policies they get. We find that the rich and middle agree more than 90 percent of the time; when they disagree, the rich win only a little more often than the middle. Even when the rich win, the resulting policies do not lean systematically in a conservative direction. Further analyses incorporating the preferences of the poorest Americans produce similar results: we find only slightly greater responsiveness to wealthier citizens. It may be that these slight differences have substantial consequences for policy and citizens, particularly as they cumulate over time; it also may be that the differences do not matter very much, and that other divisions in the electorate, for example, across partisan cohorts, or the behavior of other political actors, such as interest groups, are more relevant for inequality.

Trends in Americans' Attitudes toward Inequality

Dan Malato, *NORC at the University of Chicago*
David Sterrett, *NORC at the University of Chicago*
Jennifer Benz, *NORC at the University of Chicago*
Trevor Tompson, *NORC at the University of Chicago*

As an increasing number of policymakers identify inequality as a problem in the United States, many questions remain about the public's views of the issue. To what degree do Americans support efforts to reduce income inequality? What methods of reducing inequality do they favor and which do they oppose? How has this support changed over time? The AP-NORC Center for Public Affairs Research analyzed data from the General Social Survey (GSS) to illustrate how inequality is a complex concept that encompasses many issues, revealing the American public's nuanced views on the topic and overall

trends in attitudes regarding inequality dating back to the 1970s. We find that just under half of all Americans believe the government should reduce income differences, and the level of support for government reducing inequality has remained relatively stable for the last three decades. But, when it comes to providing assistance to lower-income Americans, the public has conflicting opinions on whether Washington should do everything possible to improve the standard of living of the poor. The data also show that Americans are sensitive to different policy approaches to reducing inequality. Differences in beliefs about inequality also emerge based on partisanship, income, age, race, and gender. The findings from this research provide insights into Americans' attitudes on a highly salient political topic and highlight the importance of capturing these attitudes through a well-rounded approach.

Income, Life Evaluation, Well-being: A Longitudinal U.S. Panel Study
Diana Liu, *The Gallup Organization*
Sangeeta Agrawal, *The Gallup Organization*

Prior research has found that income relates to well-being. According to Dr. Kahneman and Dr. Deaton's research conducted using Gallup-Healthways Well-Being Index, a daily survey of 1000 U.S. residents collected by the Gallup organization in 2008 and 2009, high income buys life satisfaction but not happiness, and low income is associated both with low life evaluation and low emotional well-being. More than five years after the worst time of financial crisis, this research uses the recent three waves (2013-2015) of Gallup U.S. longitudinal Panel data to exam how income relates to life evaluation and emotional well-being through five well-being elements (Career, Social, Financial, Physical and Community) based on a longitudinal multiple mediation model. Researchers use model fit statistics to test a predictive model and calculate overall and individual mediated effects to explain mediation through each wellbeing element after accounting for demographic differences and other influencing factors. In this study, we also use income to predict emotions and life evaluation, controlling for demographic information, weekdays or weekend, healthy behaviors and health conditions. Additionally, we conduct a causal analysis between income and emotional well-being to test whether income affects emotions and overall life evaluation.

Job Approval and Economic Evaluations: Survey Experiment Data from Wisconsin
Sarah Niebler, *Dickinson College*

Studies analyzing public opinion data have long-documented the link between partisanship and evaluations of the economy, finding that even "objective" facts such as the unemployment rate appear to be affected by partisanship. However, to what extent does the effect of partisanship on economic evaluations respond to shifts in these objective economic indicators as well as to news coverage about the economy? To answer this question, this paper uses data from the Marquette Law School Poll beginning in 2012. The MLSP contains a survey experiment repeated throughout a series of cross-sectional surveys wherein half the respondents were asked questions about President Obama's and Governor Walker's job approval before being asked to evaluate the condition of the national and state economy, while the other half were asked first about the economy and then about their evaluations of the president and governor. In this paper, I examine the degree to which the priming of partisanship (i.e. asking about job approval first) affected individuals' assessments of the economy. I also ask whether such priming of partisanship is sensitive to national and local unemployment rates as well as media coverage of economic news.

The 2016 Election and Impact on American Foreign Policy

Panel organizer: Floyd Ciruli, *Crossley Center for Public Opinion Research*

A panel will examine the impact that the 2016 US Presidential election is having on American foreign policy. The topics will include: partianship's impact on voter positions on key foreign policy issues; internationalism and militarism in a new world of threats; the Presidential debates and polarized policy positions. Additional panel members will include Kathy Frankovic, formerly of CBS News; Robert Shapiro, professor at Columbia University; and Dina Smeltz, Chicago Council on Global Affairs.

U.S. Foreign Policy is Moving to the Right. All Rhetoric or Boots on the Ground?
Floyd Ciruli, *Crossley Center for Public Opinion Research*

The 2016 U.S. presidential nomination contest with a plethora of debates and polls will be used to describe the likely trajectory of foreign policy in the next administration in several key areas. An overlay with known U.S. opinion will be integrated to identify alignment and disagreement. Discussion will include the likely influences public opinion will have on policy implementation in the next administration.

Increasing Ideological and Partisan Conflict in Public Opinion
Robert Shapiro, *Columbia University*

This presentation will show in historical perspective how ideological and partisan conflict in American public opinion has increased in the last decades. This conflict has reflected increasing polarization among Republican and Democratic leaders on a stunningly wide and far-reaching range of both foreign and domestic policy issues and perceptions of national problems -- and to a degree not seen since opinion polling began in the United States. This has important consequences for American politics and policymaking, and how we think about the role of public opinion. How long this conflict will persist is an important question for the future of the nation.

America Divided: A Growing Rift Between and Within the Parties on Foreign Policy
Dina Smeltz, *Chicago Council on Global Affairs*

With the world seeming lurching from one international crisis to the next, foreign policy will not doubt play an important role in the 2016 presidential campaign. So it is no wonder that candidates are appealing to their base voters on foreign policy issues such as immigration, terrorism and climate change. But when competing for votes next year in the presidential election, candidates will have to appeal to the median voter, which may be at odds with their base. This presentation will explore partisan differences in Chicago Council's 40 years of surveys of Americans on foreign policy and demonstrate that partisan differences on these issues have not always been as wide as they are today.

Developing a Conservative Foreign Policy
Kathy Frankovic, *CBS News*

Many self-described conservatives today favor an interventionist foreign policy. While this year's GOP candidates take a variety of positions on the role of the United States in the Middle East,

there is consistently more support for military action of all sorts among conservatives than among other groups in the population. The paper will examine the differences in positions of the major GOP candidates and their supporters on questions of terrorism and foreign policy.

Using Wearable Technology to Measure Health Outcome Indicators: A New Frontier in Data Collection

Ebo Dawson-Andoh, *Mathematica Policy Research*
Nicholas Beyler, *Mathematica Policy Research*

The rapidly expanding wearable technology market has provided consumers, universities, health institutions and private research organizations with new and readily accessible tools for measuring individual and population-level health indicators, such as physical activity, energy expenditure, fitness level, and, with some tools, even food and nutrient intakes. To date, accelerometers, pedometers, and heart rate monitors remain the tools of choice for measuring these kinds of outcomes in research settings. For example, Mathematica Policy Research recently conducted the Randomized Experiment of Playworks study, which examined the impact of the Playworks program, which places full-time coaches in lower-income schools to encourage play and cooperation, on physical activity and other school and student outcomes. Actigraph GT3X accelerometers were used to measure students' physical activity during recess and throughout the school day. To the best of our knowledge, there have been few studies that have used other forms of wearable technology for large-scale data collection efforts. We have conducted a review of the literature on wearable monitors and present a discussion on the implications for their use in future health and behavior studies may find use for these devices as part of their data collection efforts.

Perceptions about Scientific Agreement, Trust in Scientists and the American Public's Beliefs of Global Warming

Soohee Kim, *Stanford University*
Jon A. Krosnick, *Stanford University*

Prior research suggests the Americans incorrectly believe that there is much disagreement among scientists about the existence of global warming and that this 'misperception' leads them less likely to believe that global warming is occurring. However, little research has empirically examined the relationship and identified the variables that may affect the relationship. Noting that a great deal of research has demonstrated the critical role of trust in scientists for public opinion toward environmental issues, this study examines if and how trust in scientists may moderate the impact of perceptions about scientific agreement on people's own beliefs of global warming. Using data from the nationally representative sample of U.S. adults from the 2010, 2012 and 2013 surveys, this study finds that perceptions about scientific agreement was positively associated with people's own global warming beliefs, but the relationship was greater among those who have high trust in scientists. Furthermore, the finding demonstrates the importance of measurement for a better understanding of public opinion on global warming. Specifically, the finding shows the conclusion about whether or not Americans understand scientific consensus about global warming could be different depending on the question formats used in the survey: While the use of 'closed-ended (dichotomous) question,' which is the

conventional question format, leads to the conclusion that people perceive "a lot of disagreement" among scientists, the use of a new measure, 'open-ended continuous question' suggests that many people believe at least a majority of scientists (more than 50%) agree that global warming is happening. The implications of findings are discussed in terms of the importance of perceived agreement among the information source in the domain of science communication and the impact of measurement for a better representation – and therefore a better understanding of – public opinion on environmental issues.

Unbiasing Information Search and Processing through Personal and Social Identity Mechanisms
Benjamin A. Lyons, *Southern Illinois University*

Social identity models of partisanship show that voters engage in selective exposure and biased processing of information, which have deleterious effects on public opinion, vote choice and other outcomes. A considerable body of research has detailed the mechanisms of identity-protective cognition (e.g., Kahan, 2011) and resulting gaps in beliefs (Hindman, 2009; Veenstra, Hossain & Lyons, 2014), but many studies point to the difficulty of corrections (Nyhan & Reifler, 2010). Looking at more preventative mechanisms, Nyhan and Reifler (2013) show that affirming partisans' personal values can buffer against the ego-hit incurred from challenging information, reducing defensiveness. An overlooked social identity property — complexity — may offer an alternative means of mitigation (Roccas & Brewer, 2002). Whereas affirmation of personal values may reduce bias by increasing self-esteem or clarifying individuals' self-concept, social identity complexity may do so by reducing the prominence of any one identity or reducing the strictness of perceived in-group/out-group boundaries. This experiment prompted social identity complexity, social identity threat, self-affirmation, or self-derogation through a writing prompt. Respondents then selected media headlines about a fictionalized mining spill. After reading a simulated news article containing four claims blaming Republican lawmakers and the EPA, respectively (one reasonable, one conspiratorial claim against each side), they rated their agreement with the claims' validity. Results indicate that felt complexity reduced selective exposure, did not affect selective belief, but increased conspiracy beliefs regardless of target. Felt affirmation did not significantly influence selective exposure or belief, but also increased acceptance of conspiracy claims regardless of target. Self-concept clarity, however, reduced selective belief and conspiracy beliefs. These findings shed light on the mediators through which both personal and social identity influence information search and processing. In particular, complexity may "open the political mind" — but perhaps too much.

Do Facts Matter? How Political Opinions Shape Perceptions of Fact on Politicized Issues
Sarah Kelley, *University of California Berkeley*

While there have always been political debates centered on differing values, current political debates are so intransigent in part because participants cannot agree on the facts. Three important recent examples are the debates over climate change, gay marriage and the minimum wage. Are humans causing climate change? Are gay people born gay or is it a choice? Will increasing the minimum wage hurt the economy or help it? These questions - with different perceptions of fact on each side of the political aisle - seem to underlie the different policy conclusions that are reached. Yet it remains unclear whether different perceptions of these facts leads to different conclusions on policy issues or whether people's pre-existing political positions influence their perception of fact. In other words, do people who oppose environmental regulations therefore decide that humans are not causing global

warming? I approach this issue quantitatively by using structural equation modelling with appropriate instrumental variables to examine the reciprocal causation between perceptions of fact (are people born homosexual?) and attitudes about policy (should gay marriage be legal?). I find that in all three examples - the debate on minimum wage, gay marriage, and climate change - choices about policy decisions entirely shape the perceptions of facts and the perception of facts has no direct effect on policy decisions. Those who support increasing the minimum wage think it will damage the economy less (standardized effect of .40), those who support environmental regulation are more likely to see man-made global warming (standardized effect of .28), and those who support gay marriage believe more strongly that homosexuality is inborn (standardized effect of .35). This shows that far from being completely rational decision makers, people decide what is true based on what policy conclusions appeal to them, which has far-reaching implications for policy and research.

Estimating Population Size Using the Network Scale-up Method

Patrick Habecker, *University of Nebraska-Lincoln*
Kirk Dombrowski, *University of Nebraska-Lincoln*
Bilal Khan, *University of Nebraska-Lincoln*

Population sizes are a fundamental part of research, policy, and legislative agendas, but can be difficult to enumerate, particularly for hidden or hard to count groups. Considerable effort is thus devoted to developing the means and methods necessary for such enumeration. Unfortunately the resources needed to accurately measure size may place researchers in a catch-22 where in order to obtain those resources they must first know something about the size of the problem they are trying to measure. The network scale-up method (NSUM) offers a way for researchers to obtain a size estimate of a variety of groups, including hidden and hard-to-reach populations, using minimal resources and conventional sampling frames. Instead of trying to directly count the number of people in a target group in the sample, the NSUM asks sample participants how many people they know who are in the target group. When paired with a measure of the participant's personal network size for known populations, it becomes possible to generate reasonable estimates of a range of subgroups. This method was initially applied to estimates of populations such as heroin users, men who have sex with men, and those who are at risk for HIV/AIDS. However, the technique is readily applicable to the estimation of almost any group size, may be more appropriate for measuring non-stigmatized populations, and of great interest to public opinion researchers. This paper introduces the NSUM method, the mechanics required to derive estimates, error concerns, and how the method may be tailored to measure a variety of social phenomena. Finally, the paper introduces two recent improvements to the core estimation method and the treatment of back-estimates and how their inclusion improved results on a recent NSUM survey.

Response Order Effects on a Web Survey of Nurse Practitioners

Gwendolyn Quintana, *University of Texas*
Lydia E. Riley, *American Association of Nurse Practitioners*

While the purpose of this survey was to assess nurse practitioners' (NP) opinions on the social determinants of health, this poster will analyze response order effects of a national web survey of NPs. Although the subject of response order effects has been previously studied in detail, it is unclear whether response order will produce recency or primary effects in a sample of NPs, who are on average, more educated than the general population and perhaps exhibit different behaviors while taking surveys. In the summer of 2015, a random sample of 30,000 NPs were pulled from the 2015 National NP Database and assigned to one of two surveys; the only difference between the two was the order in which choices were listed for questions containing standard Likert scales for agreement, and for multi-

select questions containing a long list of response options. Subjects were sent an email invitation to a web-survey, and a subsequent email reminder was sent to those who had not responded. A total of 896 individuals completed the survey, yielding a 3.0% response rate. Significant associations between the order of response options and the frequency which response options were selected were observed for five of the nineteen questions using Likert scales. For Likert scale questions, respondents showed a tendency toward recency, selecting response options that appeared toward the end of the list. For multi-select questions with a long list of response options, respondents tended toward primacy, selecting options that appeared toward the beginning of the list. In order to minimize bias introduced by response order on web surveys of NPs, the order of Likert scales should be randomly reversed. Likewise, long lists of multi-select options should be scrambled.

Exploring Mode Effects Between Smartphone and Personal Computer Mode of Administration of a National Household Study

Kristine Fahrney Wiant, *RTI International*
Ashley Richards, *RTI International*
Stephanie Zimmer, *RTI International*
Danielle Mayclin, *U.S. Department of Energy*

Over 60% of U.S. adults now own smartphones (Pew Research Center, 2015), and smartphones are increasingly being used to respond to web surveys (Dillman 2015). Even with surveys that have been optimized for smartphone use, visual presentation of survey items on smart phones range from similar to very different than the presentation of the same items on a personal computer (PC), which has the potential to lead to mode effects. In this presentation, we examine visual layout symmetry and mode effects between smartphone and PC-based versions of a web survey, drawing on two web successive web implementations of the Residential Energy Consumption Survey (RECS). In both studies, the visual layout of the web instrument was designed to be similar across web and paper modes. Due to software constraints, however, visual differences between smartphone and PC presentations were more pronounced for question items (e.g. where a visual aid such as a photo was necessary) than others (e.g. standard radio button questions). Respondents were encouraged to respond via web in both studies, but free to response via an electronic device of their choosing. To understand any impact of device on response quality, we analyze differences in key survey estimates and item nonresponse rates between smartphone and PC modes, controlling for demographic differences related to mode selection. We classify formatting differences between smartphone and PC, and we report how these are associated with differences in survey response.

"Mental Retardation" vs. "Intellectual Disability": Differences in Responses

Marina Stavrakantonaki, *University of Illinois at Chicago*
Timothy P. Johnson, *University of Illinois at Chicago*

This poster will present findings from our investigation of the effects of a change in terminology on self-reports of an important mental health measure. This case focuses specifically on a change from use of the term "mental retardation" to "intellectual disability" in the National Health Interview Survey (NHIS). This wording change became effective as of 2011 after President Obama signed legislation requiring the federal government to replace the term mental retardation with intellectual disability in many areas of government in October 5, 2010. This measure, also known as Rosa's Law, erases the terms mental retardation and mentally retarded from federal health, education, and labor policy. "Intellectual disability" or "individual with an intellectual disability" were introduced in place of these

terms, which were considered to be hurtful language. Our analyses focus on changes in national prevalence estimates of this construct during the period 2011-2014, compared to 2007-2010, for the population as a whole, and also for demographic subgroups, using multivariate models designed to isolate potential trend changes that can be attributed to the introduction of the revised question wording. Our findings will evaluate the impact of the 2010 legislation on the continuing estimation of this construct.

CATI vs. CAPI: Mode Effects on Item Nonresponse for Sensitive Topics
Chariklia Hoefig, *Bundeswehr Center for Military History and Social Science*

In current research, different modes and mixed-modes are used to collect survey data. However, there is evidence from survey methodology research that empirical results differ depending on the respective mode. In this regard, item nonresponse is being investigated. Since item nonresponse is assumed to be not randomly distributed and, moreover, influenced by the survey mode, it needs further attention in terms of data quality. Thus, my research examines mode effects on item nonresponse in surveys about political attitudes towards security and defense policy, which are assumed to be sensitive topics and therefore should produce less item nonresponse when conducted with less interviewer contact. In order to examine mode effects, a large-scale method experiment was conducted in two concurrent surveys (CATI and CAPI) with probability samples of the German population. Both surveys included the same 44 items with varying content, stimuli, response labels and number of response options (split-ballot: 4-, 5- and 7-point scale). Hence, the robustness of the effect can be assessed. To analyze interaction effects of the mode with interviewees, various characteristics were documented. To get more information about the interviewee and the interview situation, the interviewer was asked to assess respondents' characteristics and behavior during the interview and the interview situation. In addition, metadata were recorded. In order to control for interaction with interviewer characteristics, socio-demographics and job experience were included in the dataset. Preliminary results confirm the assumption that the greater presence of the interviewer in CAPI leads to more item nonresponse in surveys dealing with sensitive issues. The mode effect is robust across different item types and interview situations and shows interaction effects with characteristics of the respondents and partly of interviewers. Further analysis suggests that the observed effects do not (only) occur due to selection bias, but are in fact a consequence of the mode.

Disagreeing in New Settings: Considering the Effects of Meme Format and Perceived Network Homophily in Online Political Talk
Benjamin A. Lyons, *Southern Illinois University*

Political talk in online spaces is governed by a new set of considerations yet to be fully articulated. First, the communicative forms have expanded, becoming more visual and viral, as exemplified in political memes (e.g., da Silva & Garcia, 2012). These have become an undeniable part of what is to talk politics online, but our knowledge of perceptions of and responses to memes is limited. Second, political talk now occurs in such a way that it is enduring across time and displayed to the breadth of individuals' networks instead of select audiences (i.e., context collapse). Awareness of these networks may also influence deliberative perceptions and behaviors. This experiment explores third-person perception and expressive responses by addressing changes in message format (meme vs. article) and implicit "others" (high vs. low perceived network homophily). In addition, this study suggests that self-affirmation – a supposed mechanism of bias reduction – can impact perceptions of media's influence on self and others. Results indicate oppositional content, network homophily, and affirmation all significantly affect third-person perception and in some cases willingness to share disagreement.

While memes did not inspire greater third-person effects than more traditional partisan media, oppositional content did across format. Meanwhile, self threat increased perception bias while affirmation diminished it. Network homophily was correlated with reduced third-person perception bias, but also with greater willingness to share disagreement, suggesting spiral of silence-type effects may hold for these online discussion scenarios. That is, those in self-homophilous networks are emboldened to share their ascendant opinions and those in self-heterophilous networks are discouraged from voicing counter-opinions. The results enrich our understanding of online talk in complex communication environments, while providing more detail about the mechanisms of perception bias.

An Adjusted Estimator for the Fay-Herriot Log-Level Model with Measurement Error in Covariates
Sepideh Mosaferi, *University of Maryland, JPSM*

The Fay-Herriot (FH) area level model in a logarithmic scale routinely used to achieve normality faster and produce reliable estimates for fund allocation and regional planning.This model can borrow strength from relevant auxiliary information to provide more precise estimates of population parameters with reduced mean squared error (MSE) compared to the direct estimator. However, if covariates suffer from measurement error, the empirical best predictor derived under the assumption of no measurement error in the covariate may result in larger MSE than simple direct estimator. In this paper, we propose an adjusted empirical Bayes (EB) and hierarchical Bayes (HB) predictor to rectify the above-mentioned problem by taking into account the MSE of the auxiliary information in the model. We illustrate the performance of our predictors through a simulation study on the 2012 U.S.Census of Governments. The results show that the proposed EB and HB predictors can lead to significant gain in efficiency over the direct estimator.

An Evaluation Study of the Use of Paradata to Enhance Data Quality
Ana Lucía Córdova Cazar, *University of Nebraska, Lincoln*
Robert F. Belli, *University of Nebraska-Lincoln*

Over the past decade, researchers have increasingly turned to the analysis of paradata to better understand and model data quality (Kreuter, 2013). In particular, it has been argued that paradata, data about the data collection process, may serve as a proxy of the respondents' cognitive response process, thereby aiding in the identification of breakdowns in surveys. The growing availability and amounts of paradata make it necessary to identify which kinds of paradata are useful in understanding these underlying processes and which are not. For instance, Cordova Cazar and Belli (2015) found that only five out of fifteen observed paradata variables had significant and practical implications for the evaluation of data quality in the 2010 American Time Use Survey (ATUS). In an effort to further extend this prior research, the following paper will focus on the investigation of changes in reported activities in the ATUS, one important aspect of the respondent-interviewer interaction hypothesized to increase the likelihood of error and reduce data quality. We compare two different operationalizations of change: an automatically generated indicator of changes in activities and a manually coded indicator of activity change, both based on the original open-ended responses. This paper systematically compares both coding schemes in terms of accuracy and explanatory power. Preliminary analyses suggest that over 35% of the automatically coded changes turn out to be substantively meaningless (e.g., due to correcting typos) and do not imply a real change in the activity when compared to the manual codings. Our results also suggest that the manually coded paradata variable is more informative and is significantly related to two types of error occurring in the ATUS (providing insufficient detail and having memory gaps). We will

discuss implications of our findings with respect to data quality, costs of coding, and strategies to improve automated coding.

A Closer Look at Response Time Outliers in Online Surveys Using Paradata Survey Focus
Stephan Schlosser, *University of Göttingen*
Jan Karem Höhne, *University of Göttingen*

Web surveys are commonly used for data collection in empirical social research because they are cheaper, faster, and simpler to conduct than other survey modes. Furthermore, they enable researchers to capture a variety of additional data (so-called paradata) during the survey process such as response times. Measuring response times has by now a long tradition in social psychological research as well as survey research to investigate response behavior and response processes. One key problem, however, is the determination of appropriate thresholds to define outliers; to a certain degree researchers determine them arbitrarily. Until now, there is no scientific consensus with respect to the definition of outliers. In our study we developed an (innovative) two-stage outlier definition procedure for web surveys using paradata. This approach is based on the activity of the web survey while processing, accompanied by an outlier definition that is based on the distribution of the response times. Our web survey (n = 1899) is based on an onomastic sampling approach and contained individual questions as well as grid questions. Moreover, we tested different procedures for dealing with outliers based on the response time distributions. Our analyses show that common outlier definition procedures, which are based on the distributions of response times, provide insufficient results. This implies that they are frequently unable to capture respondents who leave the web survey for a short time so that the response times are biased upwards. Particularly, this circumstance can be observed for grid questions. Altogether, our findings suggest that the two-stage outlier definition procedure is superior to common methods for dealing with outliers that are only based on response time distributions.

Men's Attitudes Towards Abortion: Are Men More Likely To Be Pro-choice or Pro-life In the U.S.?
Rachel Hassan, *Queens College, CUNY*
Holly E. Reed, *Queens College, CUNY*

Men make up the majority of the elected representatives in the U.S. and have a direct hand in the passing of new abortion laws or making changes to current ones. Men also make up about half of the voting population who elect those representatives. Therefore, it is important to know if men are more or less supportive of abortion because men can influence current or future abortion laws and rights for women. Although many studies have included the opinions of both men and women or of women alone, very few studies have been dedicated to men's opinions on the issue of abortion in the United States. The few studies that do exist are outdated. Data from the 2014 General Social Survey (GSS) and previous GSS data were used to examine if men are more pro-choice or pro-life, the characteristics that account for their different opinions, and how male views on this topic have changed over time. As demographic and socioeconomic factors are major determinants of people's attitudes and opinions, this research specifically focused on: race, age, education, region of residence, political party affiliation, religious affiliation, and frequency of religious service attendance as independent variables. When the seven abortion opinion questions that respondents were asked were pooled together, it was found that men were more pro-choice than pro-life. Men who were white, aged 46-60, educated

(associates degree or higher), lived in the West region, had no religious affiliation, and who attended religious services infrequently (nearly every week or less often) were more likely to be pro-choice. Men of other races (nonwhite and nonblack), who were aged 18-30, had a high school diploma or less, lived in the South region, were Republican, and attended religious services frequently (every week or more) were more likely to be pro-life. When the abortion questions were again pooled together, it was found that men's pro-choice stance on abortion has decreased over the past 40 years.

Measuring Punitiveness: Public Attitudes Towards the Punishment of Juvenile and Adult Offenders

Eva Aizpurua, *University of Castilla La Mancha*

The study of punitive attitudes has increased in recent decades. The measurement instruments to gauge them, however, have not increased in parallel. The aim of this work has been to develop and validate a scale on a representative sample of the Spanish population (n=1,000, 51.1% women, average age = 46.3). To this end a cross-validation design has been employed, randomly assigning the sample into two groups, to which exploratory and confirmatory factorial analyses were applied. The results provide evidence on the reliability and validity of the scale, composed of two first-order factors ("attitudes towards delinquency" and "attitudes towards juvenile delinquency") and one second-order factor. The differences between subscales are significant, with demands for severity greater when offenders are adults. Moreover, cluster analysis indicates that most of the subjects do not support or reject all punitive measures, but rather adopt intermediate positions.

Working with the SIPP-EHC Audit Trails: Sequential and Parallel Retrieval

Jinyoung Lee, *University of Nebraska–Lincoln*
Benjamin C. Seloske, *University of Nebraska–Lincoln*
Ana Lucía Córdova Cazar, *University of Nebraska–Lincoln*
Adam Eck, *University of Nebraska–Lincoln*
Robert F. Belli, *University of Nebraska–Lincoln*

The Event History Calendar (EHC) is a flexible interviewing tool to collect retrospective behavioral reports. The EHC method encourages respondents to use their own life events as retrieval cues, as autobiographical memories are thematically and temporally structured within and across interrelated life events. The EHC makes use of the autobiographical memory structure by promoting sequential and parallel retrieval of past events (Belli, 1998). Research has found that EHCs tend to yield higher quality retrospective reports for certain types of events or more complicated retrieval tasks, compared to conventional questionnaire instruments (Belli, Smith, Andreski, & Agrawal, 2007; Belli, Bilgen, & Baghal, 2013). The Survey of Income and Program Participation (SIPP) is a longitudinal survey of U.S. households to collect monthly information on income dynamics and governmental assistance program use. The Census Bureau has evaluated the use of EHC for the SIPP through field tests (SIPP-EHC) to reengineer the survey design. The 2014 SIPP Panel data collection began to integrate the EHC into the production SIPP. The reengineered SIPP instrument is programmed in Blaise and C#, creating audit trails during survey interviews. Using audit trails collected during the SIPP-EHC field test in 2012, this study suggests an approach for working with the SIPP-EHC audit trails that is in accordance with the model of autobiographical memory structure. Specifically, the study proposes data management strategies to parse the unstructured audit trails in a manner that enables researchers to identify the sequential and parallel retrieval processes that exist in the EHC instrument.

Consequences of Dropping the Landline Sample from RDD Surveys: Analysis from Four Years of Public Opinion Data

Courtney Kennedy, *Pew Research Center*
Kyley McGeeney, *Pew Research Center*
Scott Keeter, *Pew Research Center*
Andrew Perrin, *Pew Research Center*
Eileen Patten, *Pew Research Center*
Amanda Lee, *Pew Research Center*
Jonathan Best, *Princeton Survey Research Associates International*

Now that cell phone penetration among US adults is over 90%, many telephone survey researchers are considering just sampling cellphone numbers and forgoing a companion sample of landline numbers. Indeed, at least one major U.S. survey program, the University of Michigan Survey of Consumers, recently made that design change. Others are likely to follow, considering that a single frame design presumably entails less logistical complexity and less variance in weights (by removing the frame overlap adjustment). A central question with this looming industry-wide transition is what effect, if any, dropping the landline sample has on weighted estimates. The possibility that this design change may alter estimates is particularly worrisome to researchers studying trend estimates from telephone surveys. It is also a concern for researchers who study subgroups, such as the elderly, with lower cellphone penetration rates. In anticipation of this transition, the Pew Research Center has been generating and publically releasing an experimental weight simulating a cell-sample-only design in its monthly public opinion surveys since 2012. This presentation reports the finding of a comprehensive analysis of those data. We report the effect of simulating this transition on full sample estimates as well as estimates for subsamples defined by major demographic variables (e.g., age, gender, race, ethnicity, and education). We also examine the consequences for precision. While the change in weighted estimates associated with dropping the landline sample is generally ignorable for full sample analysis, there are noticeable differences for some subgroups and for certain topics. With respect to precision, dropping the landline sample consistently results in a lower design effect from weighting for full sample estimates. Implications of these findings for survey researchers in the U.S. are considered.

An Evaluation of Statewide Dual Frame Sample Designs in Idaho

Monica A. Reyna, *University of Idaho*
Snehalatha Gantla, *University of Idaho*

In an effort to adapt to increasing rates of cell phone only households, public opinion researchers have been incorporating cell phone sample to what were previously landline only sample designs. The state of Idaho has one of the largest estimates of cell phone only households at 52.3%. The Social Science Research Unit at the University of Idaho has completed several statewide dual-frame (cell and landline) studies in recent times. Compared to landline samples, cell phone samples tend to have higher rates of disconnects, are costly to call, and yield lower response rates. However, they have become a necessity as the prevalence of landline only households continues to decrease. Researchers must balance inclusion of land and cell phone records to achieve cost and time efficient studies while reducing non-response and coverage error. In this paper we review the development and implementation of cell phone to landline sample proportions, evaluate key statistics focused on coverage, sample performance, and cost of purchased samples, and highlight the benefits and limitations of our approach to dual-frame samples. The insight gained from this paper can contribute to

current knowledge, guide development and implementation of standards for other research units, and advance our understanding of how shifts in cell phone usage impact public opinion research.

Characterizing Satisficers in Web Surveys Using Paradata to Target Interventions

Sarah S. Vetting, *U.S. Census Bureau*
Rachel Horwitz, *U.S. Census Bureau*
Rachel Bray, *U.S. Census Bureau*
Adriana Hernandez Viver, *U.S. Census Bureau*
Jennifer Tancreto, *U.S. Census Bureau*
Courtney Reiser, *U.S. Census Bureau*

Surveys are increasingly offering web as part of a mixed mode design to elicit survey data from respondents. One concern with this approach is that respondents, in the absence of an interviewer, must be sufficiently motivated to provide accurate, complete, and honest data. Respondents that do not put forth the effort to ensure the accuracy of their responses engage in what survey researchers have dubbed satisficing behavior. Satisficing behavior can lead to data quality issues. Past studies have shown the utility of incorporating both proactive and reactive interventions into survey designs to prevent or discourage this behavior (Conrad et al., 2011; Zhang, 2013). In this study, we use the 2013 National Survey of College Graduates (NSCG) web paradata to classify respondents based on their tendency to satisfice. Specifically, we use factors such as response time, non-differentiation among answers, omissions, and early termination to identify degrees of satisficing. We also look for correlations between satisficing behaviors and demographic characteristics to help characterize these respondents. Successful classification is instrumental in targeting interventions planned for future studies based on the type of satisficing. Conrad, F. G., Tourangeau, R., Couper, M. P., & Zhang, C. (2011). Interactive Interventions in Web Surveys Can Increase Response Accuracy. Paper presented at the annual meeting of the American Association for Public Opinion Research, Phoenix, AZ, May. Zhang: http://deepblue.lib.umich.edu/bitstream/handle/2027.42/97990/chanzh_1.pdf?sequence=1&isAllowed=y

Recruiting a Survey Sample Online: Detecting and Preventing Fraud

Derick Brown, *RTI International*
Jill A. Dever, *RTI International*
Linda B. Squiers, *RTI International*
Erik M. Augustson, *National Cancer Institute*

There are a growing number of surveys recruiting and collecting data from participants solely through online methods, without any contact via landline telephone, mail, or in person. A concern for any such survey is the assurance of effective antifraud measures such that no one person is able to enter the study multiple times. The purpose of this paper is to share the antifraud methods from an innovative study, the National Cancer Institute's SmokefreeTXT Smoking Cessation Evaluation (SFTXT); a text messaging-based smoking cessation intervention. Thus a text message enabled cell phone was required to enroll into the study and receive the intervention. This study included online recruitment with a non-probability sampling method, data collection via the internet, and monetary incentives. In SFTXT, antifraud efforts centered on detecting duplicate cell phone numbers, email addresses, and IP addresses. We developed procedures during the recruitment phase of the study, along with a post-recruitment method for detecting and eliminating multiplicity when lapses in the screener antifraud methods occurred. These fraud detection methods which can be implemented in the midst of

recruitment and after recruitment are useful for any study recruiting and implementing a study entirely online, in particular for studies without a rigorous antifraud screener process.

The Best of Both Worlds: Utilizing Best Practices From Web and Survey Design
Amanda Libman Barry, *Nielsen*
Casey Langer Tesfaye, *Nielsen*
Jenna Levy, *Nielsen*

When creating a web-based instrument, it is necessary to understand user expectations for the mode. Web design often implements forms to collect user information; however, these tools do not always follow the same structure as traditional online survey design principles. For this study, we evaluate the fusion of best practices from website and survey design to create an effective online data collection tool intended to gather information from research participants in a longitudinal panel. Utilizing both passive methods, such as eye-tracking, and active methods, such as user think-alouds and retrospective cognitive interviews, we will examine the effectiveness of design changes made to the second iteration of a panelist website. The updated website incorporated user experience findings to features such as navigation tabs, card displays, the visual presentation of response options, and edit buttons. Results will be presented from an in lab test of the updated website conducted in the fall of 2015 with both former panelist as well as the general public as participants. We will present quantitative and qualitative findings related to these changes and explore the merging of best practices across website and survey design to create an effective tool for panelists. We also will discuss the importance of iterative usability testing where designs are tested, revised and then re-tested.

Making Connections on the Internet: Online Survey Panel Communications
Amanda Libman Barry, *Nielsen*
Christina Eiginger, *Nielsen*
Lauren A. Walton, *Nielsen*
Kay Ricci, *Nielsen*

To connect with potential respondents for web based surveys, researchers have several initial contact options in their tool kit. When conducting a digital study with a convenience sample, many researchers use emails, vendor driven website traffic, or banner advertisements to turn a "click" into a survey complete. Motivating willing respondents to participate in an online convenience survey may require different communications and incentives than a traditional probability survey. Taking a closer look at banner advertisements used to recruit respondents for a large convenience panels that cover participants both in and outside of the United States; this research examines digital motivational messaging and the look and feel of advertisements to understand the underling impressions and effectiveness of current sampling strategies. Findings from seventeen in-depth interviews conducted with English-speaking participants explore thoughts on existing taglines and advertisements. Overall, six taglines and seven ad copies were tested and evaluated to understand key factors in driving potential respondents to an online panel registration website. Key findings include: participant wiliness to click on the advertisements, effective incentive messaging, and comprehension of ad copy. Overall, communicating in the digital space gives researchers tools to connect with respondents, but some messages and visual designs are perceived better than others.

Assessing Potential Bias in Respondent-driven Incident Based Data from a Web Survey of College Students

Kimberly C. Peterson, *RTI International*
Marcus E. Berzofsky, *RTI International*
Bonnie E. Shook-Sa, *RTI International*
Christopher Krebs, *RTI International*
Christine Lindquist, *RTI International*

Incident-level data collection is a useful approach when measuring events that can occur multiple times within a survey's reference period. Incident-based data allow survey researchers to analyze not just characteristics of persons but also characteristics of incidents (e.g. to assess the proportion of victimizations reported to authorities). However, asking respondents to complete detailed incident reports for all incidents experienced within the reference period may be too burdensome for persons who experienced several incidents. Therefore, survey practitioners often cap the number of incident reports required for each respondent. If reported incidents differ from those not covered in the survey instrument, then bias potentially exists because limiting the number of incidents seemingly excluded incidents with certain characteristics. The Campus Climate Survey Validation Study (CCSVS), sponsored by the Bureau of Justice Statistics and the Office of Violence Against Women, was a web-based survey administered at nine colleges that collected prevalence and incident-based information on unwanted sexual contact. The CCSVS capped the number of incident reports at three and allowed respondents to determine the order in which incident reports were completed. To assess the potential for bias, we determine whether respondents systematically ordered the reported incidents. Bias could be introduced if the incidents that do not have a completed incident report are fundamentally different (e.g., occur later in the year or are less severe) than those that were reported. We consider incident ordering based on the chronological order and severity of incidents. In addition, we assess whether respondents who were unsure of the month in which one of their incidents occurred reported those incidents in a systematic way. Our analysis found that respondents do appear to systematically order their incidents both in terms of chronological order and severity. We quantify the potential impact of this biased ordering on key victimization estimates.

Desktops, Tablets and Phones, Oh My! Device Preference for Web Based Surveys

Sari Schy, *NORC at the University of Chicago*
Alyssa Ghirardelli, *NORC at the University of Chicago*
Heather Morrison, *NORC at the University of Chicago*

Web surveys are becoming increasingly popular as an inexpensive, quick and practical mode of survey administration. Correspondingly, there is also an expanding array of devices used by respondents for completing online surveys. This poster provides an empirical look at respondents' device preferences for two different samples: 1) an address-based list sample (ABS) and 2) NORC's probability-based AmeriSpeak® Panel. Survey invitations for the sampled households were sent via two different modes based on the sample source. Respondents that were part of the ABS were mailed initial letters with the survey website and login information printed on it, whereas the AmeriSpeak® Panel respondents were sent an email with a link to the survey website. The ABS subjects that did not initially respond via web were offered the option of completing via telephone or Paper and Pencil Interview (PAPI). The AmeriSpeak® Panel respondents also had the option of having initial contact and completion via email or telephone. Among the ABS respondents that completed the survey, 40% were by web while the comparable statistic for AmeriSpeak® Panelists was 75%. This poster presents analysis of device

preference for those that completed the web instrument based on sample type. This analysis helps researchers better understand how to format their online surveys based on sample type and mode of initial contact.

To Record or Ask: Measuring Respondent Gender in RDD Surveys
Margaret Tyson, *Langer Research Associates*
Chad Kiewiet de Jonge, *Langer Research Associates*
Gary Langer, *Langer Research Associates*

Interviewers in RDD surveys often are instructed to record their impression of the respondents' gender rather than asking it explicitly. However, there is little research examining the accuracy of this method. This poster examines this question by comparing interviewer impressions with data from questions directly asking respondents their gender that were included in eight nationally representative ABC News/Washington Post surveys. We find that interviewers misclassify approximately 2 percent of respondents by gender, with more women misclassified as men then men as women. This results in a net shift in gender of approximately 1 percent on average. The paper explores impacts on weighting and examines predictors of misclassification using respondent and interviewer characteristics.

Predicting Public Reception of Scientific Evidence Using the Scientific Reasoning Scale
Caitlin Drummond, *Carnegie Mellon University*
Baruch Fischhoff, *Carnegie Mellon University*

Since the late 1970's, the National Science Foundation has sponsored the collection of data related to Americans' science literacy and attitudes toward science and technology. The science literacy items used by the NSF were developed from public understanding of science research (Miller, 2004). They test factual scientific knowledge, using true/false questions, and knowledge of probability and experimental design, using open-ended and multiple-choice questions. Our research builds on this research by combining behavioral decision research and cognitive developmental psychology to identify and then measure the skills that laypeople need to understand and interpret scientific evidence. Cognitive developmental psychologists, in studies of how individuals develop the skills used in scientific inquiry, have identified two key aspects of individual ability to perform scientific inference: scientific knowledge and scientific reasoning. Scientific reasoning ability is defined as the ability to "think like a scientist," by using the reasoning processes key to scientific inference (e.g., causal inference). We develop a psychometric measure of scientific reasoning ability, the Scientific Reasoning Scale (SRS). We assess the reliability and validity of the SRS using several small-sample surveys with convenience samples. Our results indicate that the SRS is internally consistent and distinct from the existing NSF science literacy measures. Individuals with higher SRS scores are more likely to have beliefs consistent with the scientific consensus on contentious issues such as vaccination and global warming, above and beyond education, political and religious beliefs, and scores on the NSF science literacy measures. We discuss the implications of our research for developing new indicators for public reception of scientific evidence.

Debunking Myths About the Quality of Industry and Occupation Data Collected Through Self-administered Web Surveys

Felicia I. Hurwitz, *Mathematica Policy Research*
Jillian Stein, *Mathematica Policy Research*
Amanda L. Skaff, *Mathematica Policy Research*

Currently, survey methodologists are concerned with the data quality of industry and occupation responses collected through self-administered web surveys. Such surveys are conducted without the assistance of interviewers to elicit detailed information to ensure that respondents' employment information can be successfully coded into standard industry and occupation categories. The Evaluation of the YouthBuild Program is currently collecting industry and occupation data via a multi-mode longitudinal survey. Respondents have the option of completing the survey on the phone through computer-assisted telephone interviewing or through a self-administered web option. This paper will explore the results of collecting industry and occupation data on the web by comparing the data quality of interviewer-assisted responses to that of self-administered survey responses for industry and occupation questions. Data quality will be measured by the ability to code responses using the standard codes developed by the Bureau of Labor Statistics. Our data analysis to date suggests that high quality industry and occupation data can be collected using a web survey.

The Nexus Between Data Science and Survey Design and Practice

Steven B. Cohen, *RTI International*
Gayle S. Bieler, *RTI International*

The field of data science has served to rapidly expand the knowledge base and decision-making ability through the combination of seemly disparate and diverse sources of information and content, which include survey and administrative data, social, financial and economic micro-data, and content from mobile devices, the internet and social media. Other attributes of data science include data visualization and rapid prototyping; social media analytics and social network analysis; predictive, mathematical and simulation modeling; use of Bayesian methods, machine learning; GIS and geospatial analytics and Big Data technologies. When appropriately harnessed, the resultant outputs have the capacity to serve as catalysts to yield new, actionable insights. Advances in data science also serve to facilitate the effective and efficient utilization of statistical models and procedures in concert with big data applications. Through these synergies, greater accuracy and efficiencies can be realized in the application of predictive modeling techniques, the implementation of optimization analytics, and in the development of enhanced data resources that can provide the informatics to guide decisions and target interventions. In this presentation, attention is given to demonstrate the capacity of data science to inform the design of surveys, their operations and associated strategies to reduce survey errors and enhance data quality. Examples are provided as they apply to sample frame development, sample design specifications, operationalizing oversampling strategies and analytic file creation. In each of these settings, we highlight the gains in efficiency, accuracy and quality in survey procedures realized through specific applications of data science to ongoing survey efforts.

Applying a Blended Learning Model to Field Staff Training
Deborah R. Herget, *RTI International*
Sameena Y. Siddiqui, *RTI International*
Colleen M. Spagnardi, *RTI International*

In-person field data collector training for a national study is a costly activity considering hotel, travel, and meal costs for field staff and the added cost of remedial and refresher training. To address these issues, a hybrid model of online and in-person training was introduced for two national school-based studies conducted by the National Center for Education Statistics (NCES). This hybrid model, commonly referred to as blended learning, was designed to decrease the number of in-person classroom days, improve training quality, and reduce costs by developing a combination of study-specific and generic modules that can be used over the lifespan of longitudinal studies and other related studies. Blended learning has been found to improve pedagogy, increased access and flexibility, and increase cost effectiveness (Graham, Allen, & Ure, 2003). The online training modules were made available to field data collectors about two weeks prior to the in-person training to allow staff to complete the training at their convenience. Because trainees receive much of the content beforehand, the in-person training was then designed to focus on hands-on, interactive activities. Data from quizzes at the end of each module was used to adapt the in-person training to focus on topics that needed more attention or hands-on practice. Providing content online also allowed for remedial review later. This paper will investigate the efficacy of the hybrid training model, looking at the cost and quality of training when blending online self-study and in-person instruction. Cost will be measured by comparing costs incurred for developing the online training and time for field staff to complete the online training to the cost savings realized by reducing the time and expenses for in-person training. Training quality will be measured by the need for supplemental training after the in-person training and trainee evaluations.

Keeping Up with the Times: A Web Diary Design for the Consumer Expenditure Survey
Nhien To, *U.S. Bureau of Labor Statistics*

Technology dominates fast-paced modern life. To remain relevant, survey research must adapt to the ever-changing technological landscape. Currently, the Bureau of Labor Statistics (BLS) Consumer Expenditure Survey (CE) collects data on small frequently purchased items by asking respondents to hand write all of their household members' expenses in a 44 page paper diary. In 2009, the CE program launched its multiyear Gemini Project to research and develop a redesign for the CE Survey which included a plan for developing a web diary option. By taking advantage of technological advances, particularly the portability and prevalence of mobile devices, a web diary can offer other benefits that may help to improve data collection. Individuals in a household can access and report their own expenses, reducing the burden on one household respondent and relying less on proxy reporting; respondents can record their expenses as they occur or soon after, reducing recall time; data can be entered directly, eliminating transcription errors; and respondents have the option to choose from multiple modes, catering to different preferences. In 2011, CE began the process of building a web diary accessible via a desktop computer followed by a mobile version in 2013. A series of usability and field tests were conducted, leading to a number of change recommendations and multiple design iterations. With very specific data requirements, creating a user-friendly web diary was not easy, but CE has made good progress and is nearing the completion of its web diary development. This poster presentation shows the progression of the CE web diary from conception to its current state, sharing findings and lessons learned along the way as well as plans for moving forward. The findings will be of interest to

survey methodologists working to integrate web or mobile design options into their survey data collection modes.

Who Gets It? Comprehension Difficulties in Political Survey Questions
Naomi Kamoen, *Utrecht University*
Bregje C. Holleman, *Utrecht University*

Voting Advice Applications (VAAs) have become a central source of political information during election time in many European countries. VAAs are web-based tools in which users answer survey questions about political issues and receive a voting advice based on their answers. In the current research, we employed two methods to evaluate comprehension problems in VAAs. First, we conducted cognitive interviews among 60 VAA users during the Municipal Elections in the Netherlands in 2014. Second, we gained access to all the VAA answers provided to the VAA statements during this election period, which allowed us to statistically analyze the relation between response patterns and question characteristics. Results of the cognitive interviews show that users encounter comprehension difficulties for about 1 in every 5 questions. These relate to semantic meaning problems (e.g. misunderstanding technical terms, unknown geographical locations) and pragmatic meaning problems (e.g. lack of knowledge about the issue at stake). In case of comprehension problems, VAA users fill in the gaps of missing information by making assumptions about the meaning of/about the question, rather than by looking for factual information on the web. Failing to look for additional information, VAA users nevertheless proceed by supplying a substantive answer on the 'agree'–'disagree' dimension (75% of the cases). Most of these substantive answers are neutral answers. Given that cognitive interviews indicate that, first, the usage of political terms and geographic locations causes comprehension difficulties, and second, that comprehension difficulties translate to no-opinion answers and neutral answers, we expected to find correlations between these factors in large-scale datasets. Analyses of 34 real-life VAAs (34*30 questions, over 300,000 respondents) confirmed that the mentioning of political terms and locations was correlated with larger proportions of no-opinion responses and neutral responses. Therefore, we advise VAA developers to avoid using technical terms and locations in their surveys.

Hands-free Testing: Applying Automatic Methodologies to Data Collection and Cleaning
Larry M. Vittoriano, *Mathematica Policy Research*
Sean Kirk, *Mathematica Policy Research*
Matt Potts, *Mathematica Policy Research*
Taylor Calise, *Mathematica Policy Research*

Mathematica Policy Research has developed a methodology to completely integrate the data collection and cleaning process, from designing a data collection system, testing its veracity, and consistently verifying and validating the data collected through the field period. Using transparent and comprehensive specifications, and a diverse set of automated tools, we have dramatically reduced the probability of human error in the data collection process. We achieved these gains by removing most of the human component from the process. This exciting new methodology adds a level of rigor unmatched by other methods and drastically reduces the time from survey administration to final analytic data set. Using a unified, comprehensive set of specifications, we design, program, and test the data collection instruments and data cleaning processes to ensure continuity. During the field period, we use data analytics technology to concurrently review and validate the raw data collected. The result is

an increased level of confidence in the integrity of the data via a transparent process that facilitates more efficient assessment of data anomalies and provides far more time for data analysis than more conventional approaches allow. In effect, the process lowers the costs for programming and testing efforts, resulting in decreased data collection and cleaning costs. Using various metrics on the testing process, the paper describes how these tools and techniques (1) reduce errors in data collection instruments across modes, (2) verify the accuracy of instruments in the most complex scenarios, and (3) contribute to a reduction in the total survey error.

Influence of Family-based Factors on Response Propensity in the Project Talent Twin and Sibling Study

Deanna Lyter Achorn, *American Institutes for Research*
Danielle Battle, *American Institutes for Research*
Ashley Kaiser, *American Institutes for Research*
Carol A. Prescott, *University of Southern California*

The strength of twin and family studies relies upon the cooperation of family pairs. For longitudinal family-based studies, gaining cooperation is based on two steps—first, successfully locating family members and second, successfully encouraging their response to the questionnaire. Research on the cooperation level of families has indicated that differences in health, personality, and lifestyle for families that are more or less cooperative are small (Vink, et al. 2004), but that response propensity is, to some degree, heritable (Thompson, et al. 2010). This suggests that twins and their siblings as a family set may be more or less likely to participate, but less is understood about the contextual factors within a family that could impact willingness to participate. The Project Talent Twin and Sibling Study (PTTS) includes twins, triplets and their siblings (N=5,161) who participated in Project Talent in 1960. Launched in 2013 in partnership with the University of Southern California, AIR located 96.7% of individuals and 94% of twin/triplet sets. Our analysis indicates that tracking success was primarily a function of the quality of the tracking variables that were available. Base-year indicators identifying families that may be more difficult to locate (e.g., immigrants, more mobile families) as well as personality and cognitive variables were not significantly associated with tracking success. This finding was consistent at both the individual and family levels. This analysis continues our investigation by examining what family-based factors impact individual- and family-level response propensity. Is there a relationship between the likelihood of responding and family-based factors, such as the vitality status of siblings, the presence of an early responding sibling, and indicators of close relationships with siblings? We will also take a closer look at nonrespondents, differentiating between cases with no contact and refusals, to assess the likelihood that a participant may have been unsuccessfully tracked.

Cross-cultural Measurement Invariance Among German Migrants in Welfare Benefits Receipt

Jonas F. Beste, *Institute for Employment Research (IAB)*
Arne Bethmann, *University Mannheim*

An emphasis of many surveys is the measurement of subjective indicators concerning a wide field of topics. The measurement instruments used for these purposes (e. g. batteries of multiple items) rely on the assumption of measurement invariance. This means, that all respondents have a similar understanding of the measured underlying construct as well as each individual item. To compare means of different groups of respondents we must ensure that these groups understand and respond to the questions in similar ways. Otherwise comparison between groups can lead to incorrect conclusions.

Previous methodological research has shown that measurement invariance is not given for all instruments and groups. Particularly, differences appear between groups of different cultural background (Davidov et al., 2014). Therefore testing measurement invariance is of utmost importance for surveys including respondents with cultural diversity. The Panel Study "Labour Market and Social Security" (PASS) is an ongoing yearly household panel study of German welfare benefits recipients and is concerned with their living conditions, socio-economic situation and the dynamics of welfare receipt. Culturally the PASS respondents are very heterogeneous due to the large proportion of individuals with a migration background. This is intensified by the rising number of refugees from Syria, Iraq and Afghanistan over the last month, which enter the study through annual refreshment samples. The recent developments and their socio-political implications increase the need for valid sociological insights. To assure comparability between groups of different cultural background we test multiple measurement instruments for multiple subjective indicators in PASS (e. g. well-being, life satisfaction, social participation and inclusion, psychometric measures) using a multi-group CFA framework (see Vandenberg & Lance 2000). We operationalize cultural background using questions on migration background, spoken language, religion and nationality as well as the language the interview was conducted in.

Patterns of Unit and Item Nonresponse in a Multinational Web Survey

Allison Ackermann, *Abt SRBI*
Elaine Howard Ecklund, *Rice University*
Benjamin Phillips, *Abt SRBI*
Alec Brulia, *Abt SRBI*

We examine nonresponse in the Rice University Religion among Scientists in International Context (RASIC) survey. RASIC was a web survey with telephone follow-up of members of biology and physics departments in French, Italian, Turkish, and U.S. universities and research institutes. Our analyses provide a guide for other studies of academic populations with respect to response rates for a typical web survey design and with respect to patterns of unit and item nonresponse. We first examine cross-national differences in response rates (highest in Italy and the U.S., lowest in Turkey). The timing of response is examined next, noting localized influences on the pattern of responses (e.g., holidays and the timing of reminders and telephone follow-up). This is followed by analysis of variations in response rate within country by characteristics of scientists, including academic rank, discipline, gender, and elite/nonelite status of institution. Differences between early and late respondents are then examined to allow for the use of the wider range of variables measured in the survey itself, including measures of religious belief. Finally, we examine which items were associated with break-off and item nonresponse: questions on religion, for example, could be controversial in some countries. The results of this analysis will provide data on rates and patterns of nonresponse in academic populations in the countries studied. RASIC data collection was funded by the Templeton World Charity Foundation, grant TWCF0033.AB14, Elaine Howard Ecklund, PI, Kirstin RW Matthews and Steven W. Lewis, co-PIs.

Dynamic Instructions in Check-All-That-Apply Questions

Tanja Kunz, *Darmstadt University of Technology*
Marek Fuchs, *Darmstadt University of Technology*

In check-all-that-apply questions, respondents are required to select all applicable responses. Although check-all-that-apply questions are one of the most commonly used question formats in (Web) surveys, respondents often do not spend sufficient effort to thoroughly process each of the response

alternatives. Instead, respondents tend to select one of the first alternatives without sufficiently considering the remaining ones, resulting in primacy effects and an overall lower number of responses then actually apply to them. On the contrary, respondents may select response alternatives that apply to them only vaguely, resulting in a considerably higher number of responses than desired. In order to ensure comparability researchers often use instructions with check-all-that-apply questions that specify the number of responses desired (e.g., "Please select the three most important aspects."). However, such instructions are often overlooked or ignored by respondents. Web surveys offer the opportunity to implement dynamic design features that possibly increase the respondents' attention to such instructions. In this paper, we assess the effectiveness of instant feedback messages appearing once respondents start answering a question. Using a between-subjects design, the effectiveness of providing instructions either in the form of static instructions that are always visible together with the question stem (EG1), dynamic instructions that instantly appear once respondents start answering the question (EG2), or a combination of both (EG3) was assessed. Experimental conditions were evaluated compared to a control group where no instruction was provided (CG). Initial findings concerning the effectiveness of different instruction types showed that a combination of both static and dynamic instructions is most effective in obtaining the desired number of responses. Comparisons with importance ratings shed light on the question whether respondents actually select the most important responses. In addition, response order effects were assessed in order to determine the extent of satisficing behavior in each experimental condition.

Timing is Everything: Discretely Discouraging Mobile Survey Response through the Timing of Email Contacts

Ashley Richards, *RTI International*
Amanda C. Smith, *RTI International*
Bonnie E. Shook-Sa, *RTI International*
Marcus E. Berzofsky, *RTI International*

The proportion of web survey responses submitted from mobile devices, such as smartphones, is increasing steadily. This trend is problematic because mobile responses are associated with increased breakoffs, item nonresponse, and other data quality issues. Careful web survey design can mitigate some of these concerns, but cannot eliminate them entirely. As a result, survey practitioners typically prefer that respondents not respond via mobile devices. Web surveys can be programmed to block mobile responses, but this approach is discouraged because of its potential to increase nonresponse (Buskirk & Andrus, 2012). Ideally, researchers need a way to discourage mobile response without impacting response rates. In this paper we evaluate a strategy for discretely discouraging mobile responding. The Campus Climate Survey Validation Study Pilot Test, sponsored by the Bureau of Justice Statistics and the Office on Violence Against Women, is a survey of college students at nine U.S. institutions. Over 23,000 respondents completed the survey among a random sample of approximately 50,000 students. Although schedules vary across students and institutions, we suspect college students are less likely to respond via mobile devices during certain times of the day than others. For example, in the early evening on a Monday, they may be more likely to be using a computer to complete assignments, and thus less likely to respond via mobile devices. Using data on the day and time a response was submitted as well as the day and time a respondent was last emailed a request to complete the survey, we identify the times that are most likely to result in non-mobile responses. Because web survey response typically spikes immediately after invitations and reminders are sent, the findings of our research can be used to carefully time email contacts in an attempt to discretely discourage mobile responding among a college student sample.

Great Jobs Associate with High Well-being Worldwide
James Bird, *The Gallup Organization*
Sangeeta Agrawal, *The Gallup Organization*
Diana Liu, *The Gallup Organization*

High-quality jobs keep the world moving forward. Jobs that tap the energy and creative potential of populations everywhere. Jobs that give individuals a sense of pride and dignity as they build a more secure future for themselves and their families. Jobs that allow people to apply their strengths in ways that contribute to the growth and success of businesses, industries, and even entire societies. Unfortunately, in most of the world, those types of jobs are hard to come by. A recent Gallup global great jobs report reveals that out of 5 billion adults worldwide, 1.3 billion have a good job. Of these 1.3 billion, roughly 12% are engaged. Out of a global workforce of an estimated 3.2 billion adults who are working or looking for work, only 5% or 161 million people have a great job. We raise the question of whether people who have great jobs have high well-being, and whether great jobs associate with positive emotions and well-developed economics. We report an analysis of more than 300,000 responses in 148 countries based on Gallup World Poll conducted in 2014 and 2015. This study reveals the relationship between people who are engaged at work and their life evaluation/well-being at both worldwide and regional levels, and also uses great or good jobs to predict five well-being elements (Career, Social, Financial, Physical and Community). Additionally, we report and visualize the association between people who have great jobs and important indicators in the areas of Economy & Growth, Education, Health and Emotional Well-being at the country level.

Has Historical Survey Response Error Masked the Prevalence of Mental Health Conditions in the United States?
Dianne Rucinski, *Independent Researcher*

Among the many changes brought forth by the Patient Protection and Affordable Care Act (PPACA) was the promise of preventive health services and fuller realization of mental health care parity, as envisioned in the 2008 Mental Health Parity and Addiction Equity Act. Assessing the impact of the PPACA on population mental and behavioral health and service utilization is complicated by evidence presented over the last two decades suggesting that increases in mental and behavioral health treatment has not yielded improvements in population-based mental health measures, such as reducing the prevalence of major depressive episodes in the past 12 months and severe mental illness (Kessler, et. al, 2003; 2005). Among the many reasons suggested for failures to find positive associations between treatment and lower prevalence of mental illness is that a growing acceptance of mental and behavioral health treatment and concomitant destigmatization of mental illness have brought forth higher self-reports of mental and behavioral health problems in population surveys (Mojtabai, 2007; 2009; Jorm, 2014). This position argues that serious underreporting of mental and behavioral health problems in population surveys in the past masked significantly higher prevalence rates and that current prevalence rates would even higher without increased treatment. To investigate this hypothesis we examine trends self-reports of mental health issues and substance use disorders in population surveys collected through interviewer administered surveys and self-administered surveys. It has been demonstrated that self-administered surveys measuring major depressive episodes and generalized anxiety disorder result in higher prevalence rates than interviewer administrated forms (Epstein, Barker, Kroutil, 2001). Specifically, data from the Behavioral Risk Factor Surveillance System, the National Health Interview Survey, and the National Survey on Drug Use and Health will be used to test the destigmatization and

treatment acceptance hypothesis that self-reporting of mental and behavioral health issues has changed over time.

Building a Taxonomy and Lexicon of Terms and Concepts at the U.S. Bureau of Labor Statistics

Daniel W. Gillman, *U.S. Bureau of Labor Statistics*
Elizabeth Ashack, *U.S. Bureau of Labor Statistics*
Daniel Chow, *U.S. Bureau of Labor Statistics*
Ronald Johnson, *U.S. Bureau of Labor Statistics*
Karen Kosanovich, *U.S. Bureau of Labor Statistics*
Ann Norris, *U.S. Bureau of Labor Statistics*
Garrett Schmitt, *U.S. Bureau of Labor Statistics*
Thomas Tedone, *U.S. Bureau of Labor Statistics*
Clayton Waring, *U.S. Bureau of Labor Statistics*

The taxonomy and lexicon project at the US Bureau of Labor Statistics (BLS) began in summer 2013 with the goal to provide consistent access to BLS time series data and documents. A team consisting of experts from most of the offices within BLS was formed to conduct the work. The end result is the taxonomy will support searching for data and provide guidance for website redesign, while the lexicon will support tagging, cataloging, and searching for documents. The technical language used by BLS is sometimes misunderstood by users. For instance, users searching for data using the word "Cities" might not locate data labeled with the technical term "Metropolitan Statistical Areas". The main problem is the identification and organization of all BLS technical terms into a useful terminology system. This includes identifying plain English words, such as "Cities", users often employ in their searches for BLS data. The team has worked in phases. In the first two of these, identification of plain English words was completed, and a first draft taxonomy was built. The plain English were incorporated into the titles of the categories at the top levels of the taxonomy. Additionally, the taxonomy was divided into facets, one addressing the measures BLS produces, and the other supporting the so-called characteristics, or classifications, used to stratify or sub-divide the measures. Some of the most difficult work was in organizing the measures under the higher level categories. In the third phase, the taxonomy underwent quality improvement. The results of this work were given to the program offices in BLS for review. Then, the taxonomy was submitted for cognitive testing of people outside the agency. The paper includes a detailed discussion of the work in each of the phases of development with special attention given to the quality improvement steps.

Analysis of Weighting Controls and Their Effect on Data Validity

Kalina Popova, *Nielsen*

This paper looks at the effect of weighting controls (targets), both variables included and their alternative specifications, on the accuracy of survey data. While most survey data requires weighting to assure it is representative of the population of interest, there is a variety of weighting specifications used even within the same company. We set out to evaluate what we currently use and whether we should change the variables we use, whether removing some, adding new ones, or modifying the variable specifications, with a view ultimately at the accuracy of our survey data. We identified 24 benchmarks and ten potential weighting control variables, both at the household and individual level. Then we collected non-probability survey data for comparison. The evaluation consisted of two parts. First we identified the rank order of importance of the potential weight controls on each of the

benchmark measures based on the incremental increase of the log likelihood ratio in sequential logistic regressions. Then we estimated the mean absolute error of survey responses from the benchmark measures using alternative weighting schemes. As a result we identified new variables to include in our weighting model going forward, different variables to include depending on the survey topic, and modifications to the specification of most of the variables we had been using so far. The new weighting scheme improves the accuracy of the survey data vis-à-vis benchmarks while at the same time still providing comparable weighting efficiency to the one of the previously used model.

Physical Measures, Specimen Collection and Health Apps: Public Willingness to Participate in the Next Generation of Health Surveys

John M. Boyle, *ICF International*
Lew Berman, *ICF International*
Jamie Dayton, *ICF International*
Ronaldo Iachan, *ICF International*
Eric Miller, *ICF International*
Melanie Courtright, *Research Now*
Kartik Pashupati, *Research Now*

For more than seventy years, we have measured the health of the US population primarily through population surveys using self-reported health measures. However, there is increasing pressure to include objective physical measures and specimen collection in future health surveys. Potentially, this next generation of health surveys would not only provide better data, but support new health programs like the NIH Precision Medicine Initiative. However, aside from cost and logistics, this next generation of health surveys requires public willingness to participate in additional objective physical measurement, specimen collection, and behavioral measurement. We have relatively little information on the willingness of the general public to permit interviewers to take a respondent's weight and blood pressure in their homes, do finger pricks for blood tests, or add apps to their smart phones to record physical activity. Moreover, we do not know who is willing and not willing to participate. It would be useful to profile the characteristics of those who would be willing and develop model predictors for the willingness to provide these data. We address these questions using the findings of a community health survey conducted among 1200 adults living in Los Angeles. The interview was conducted among a community sample drawn from a national web panel. At the end of the interview, survey respondents were asked about their willingness to participate in blood pressure testing and finger prick blood collection in their homes, and adding health apps to their smart phones, in order for federal and local health agencies to get better measures of community health. We will report on the willingness to participate in these measures, and the health and demographic characteristics of those respondents. This should provide some initial insights into the challenges facing the next generation of health surveys.

Comparing Distribution Protocols for Leave-behind Collection of Accelerometry in a Population Survey

Daniel Lawrence, *NORC at the University of Chicago*
Katie O'Doherty, *NORC at the University of Chicago*
Meredith Czaplewski, *NORC at the University of Chicago*
Kriston Koepp, *NORC at the University of Chicago*

It is increasingly common for large-scale surveys to include measures of physical activity. Survey researchers, scientists, and doctors are exploring ways to use the objective metrics of wearable

technology to validate or augment subjective survey responses. With the growth of these wearable devices, it is important to understand how the method of distributing such devices may shape a wide range of outcomes. NORC has now collected data on sleep and physical activity during two separate waves of the National Social Life, Health and Aging Project (NSHAP). As part of the NSHAP computer-assisted personal interview (CAPI), selected respondents are asked to wear a device called an accelerometer on their wrist for several days. The method for distributing the accelerometer differed between the two most recent waves of NSHAP. During the second wave, interviewers explained the accelerometry measure to respondents during the in-person interview. Then, another staff member called respondents to arrange a time to mail them the device. This process meant that some respondents who agreed during the in-person interview could not be reached later to receive a device. During the third wave, accelerometry data were collected using a new device which was integrated directly into the questionnaire. Interviewers again explained the measure during the in-person interview, but then activated the device with a single-click in the CAPI instrument and helped the respondent put on the device immediately. This paper describes these different approaches in detail and provides cross-wave comparisons of our cooperation rates, return rates, number of prompts, average time to complete the measure, timely or delayed returns of the materials, respondent characteristics, and methods to ensure data quality. We will discuss how the lessons learned from implementing different distribution protocols can inform other studies considering the use of similar measures.

Statistics Canada's Experiences in Implementing and Managing Responsive Collection Design for CATI Surveys

Francois Laflamme, *Statistics Canada*
Lise Rivais, *Statistics Canada*
Sylvie Bonhomme, *Statistics Canada*

Over the past decade, paradata research has focused on identifying strategic data collection improvement opportunities that could be operationally viable and lead to improvements in quality or cost efficiency. To that extent, Statistics Canada has developed and implemented a Responsive Collection Design (RCD) strategy for Computer-Assisted Telephone Interview (CATI) household surveys to maximize quality and potentially reduce costs. RCD is an adaptive approach to survey data collection that uses information available prior to and during data collection to adjust the collection strategy for the remaining in-progress cases. In practice, the survey managers monitor and analyze collection progress against a pre-determined set of indicators for two purposes: to identify critical data collection milestones that require significant changes to the collection approach and to adjust collection strategies to make the most efficient use of remaining available resources. In the RCD context, numerous considerations come into play when determining which aspects of data collection to adjust and how to adjust them. Paradata sources play a key role in the planning, development and implementation of active management for RCD. Since 2009, Statistics has conducted several RCD surveys. This paper describes Statistics Canada's experiences in implementing, and especially, in monitoring these surveys. In particular, this paper presents the plans, tools and strategies used to actively manage the RCD surveys and how these strategies evolved and improved over time.

Developing and Testing Items to Measure Limited English Proficiency for the MCBS

Rene Bautista, *NORC at the University of Chicago*
Lisa Lee, *NORC at the University of Chicago*
Samuel C. Haffer, *U.S. Centers for Medicare & Medicaid Services*
Paul Guerino, *U.S. Centers for Medicare & Medicaid Services*

Elderly and disabled Americans with limited English proficiency (LEP) may be particularly vulnerable to decreased access to and satisfaction with health care and poorer health outcomes compared to those who are proficient in English. In an effort to reduce these disparities, it is essential to accurately identify the LEP status and preferred language of a health care recipient as required by the Department of Health and Human Services (DHHS) data collection standards, and to evaluate measures addressing how LEP affects the receipt and comprehension of health care information from health care providers. To address these issues, NORC, in collaboration with the Centers for Medicare & Medicaid Services (CMS), developed and is testing a new set of LEP measures for the Medicare Current Beneficiary Survey (MCBS). The MCBS is a nationally representative survey of the Medicare beneficiary population, which collects extensive information on health care use and expenditures, sources of and access to health care, as well as satisfaction with care. The goal of this paper is to assess how beneficiaries with LEP can be identified and to assess their receipt and comprehension of health care in a large scale health survey. This LEP module, which queries Medicare beneficiaries about their proficiency with spoken and written English and communication with their health care provider, is being cognitively tested for the MCBS. The presentation will include results for approximately 30 cognitive interviews with Medicare beneficiaries recruited through organizations serving the elderly population in the Chicago area, some of which were conducted in Spanish and some conducted with the help of a language assistant. We will present the results of the cognitive testing, the process for developing the LEP module, as well as issues related to the use of language assistants in the cognitive interviews.

Increasing Motivation, Engagement and Data Quality Through a Child-friendly Online Platform for Survey Data Collection with Children

Melissa DeRosier, *3C Institute*
Janey McMillen, *3C Institute*

Within clinical, social, and behavioral fields of study, survey data collection is traditionally carried out in-person, by telephone, or through the mail and often involves research participants completing one or more questionnaires via paper-and-pencil forms. While evidence supports significant advantages of online data collection, such as higher response rates and substantial cost savings, none of the current software products offer developmentally appropriate tools specifically designed for children. The same platform for survey delivery is applied regardless of respondent age, with only minor accommodations for younger respondents (such as graphic smiley face response choice options). While adolescents may be able to respond equally well to an adult formatted questionnaire, developmental challenges in literacy, motivation, and attention span can significantly alter the quality of data collected with children younger than 13 years. In this presentation, observational data, software metrics, and survey results (e.g., psychometric reliability and validity indicators) gathered through a study of 60 K-5th grade children with and without learning disabilities will be presented. In this study, children completed parallel survey items via (a) traditional paper-and-pencil group administration methods (with read aloud by the researcher) and (b) a child-friendly gamified online data collection platform with built in voice-over and developmental accommodations (e.g., audio synchronized highlighting). Discussion of results

will focus on the ways by which a child-friendly data collection platform can be used to effectively increase motivation, engagement, understanding, and data quality for survey responses by children. We will also discuss the relative value of specific developmental accommodations when collecting survey data with children with and without learning disabilities.

Surveying Rural Appalachia: A Study of Modes, Response Rates and Cost in a Rural, Hard to Engage Population

Morgan Jones, *East Tennessee State University*
Kelly N. Foster, *East Tennessee State Unversity*

Due to declining response rates, one of the biggest concerns among researchers is how to get the best response rate at the most reasonable cost. This is particularly a problem in rural communities, which have the same concerns that we see nationally, but also have high levels of distrust, literacy issues, and geographic limitations. The purpose of this study is to examine the most effective strategies for engaging potential respondents among a rural, hard to engage population. In fall 2015, the Applied Social Research Lab (ASRL) conducted a Community Needs Assessment for the town of Jonesborough, Tennessee. Potential respondents were initially contacted via USPS with a standard invitation packet. The response rate was only 3%. Because of this low response rate, we were concerned that additional contacts via USPS would not yield the number of responses desired. In order to determine the most efficient mode (highest response and lowest cost) we randomly split 3,000 of the non-responser's addresses into three follow-up strategies: phone call, USPS mail packet, and in-person visit. Respondents who were in the phone contact group were significantly more likely to complete the survey than the mail and face-to-face groups[1]. Telephone was also the most cost effective, with the cost per complete nearly 5 times cheaper than mail and 19 times cheaper than face-to-face. The most significant contribution of this research is that it explores how successful traditional methods and contact modes are in a rural, hard to engage populations, as well as how the response rates differ from other regions of the country. While nationally there is a concern over the future viability of telephone surveys, this may not be warranted in rural, Appalachian communities. In these communities the telephone may be the most stable and efficient mode of survey administration. [1] X2 (2) = 40.939, p<.05.

RDD Telephone Sample Trends: Improving Efficiency in the BRFSS

Andrew M. Evans, *Abt SRBI, Inc.*
Dennis Daly, *Abt SRBI, Inc.*
Fred Breukelman, *Delaware Division of Public Health*
Elizabeth J. Ferree, *Virginia Department of Health*
Jillian Garratt, *Ohio Department of Health*
Kenneth J. O'Dowd, *New Jersey Department of Health*
Madhavi Vajani, *Georgia Department of Public Health*

As consumer telephone usage behavior has shifted away from landline phones and towards greater reliance on mobile phones over the past decade, survey practitioners have increasingly struggled to balance the higher costs of cell phone interviewing against the declining representativeness of the landline frame. More recently, the landline frame has seen decreasing efficiency related to both increased rates of nonworking numbers and declining response. Abt SRBI currently conducts the Behavioral Risk Factor Surveillance Survey (BRFSS) in 6 states. The CDC has made several changes in the BRFSS sample protocol in 2015 to improve the efficiency of calling both landline and cell phones. These changes include removing cell phone numbers flagged as inactive through a prescreening process by the

sample vendor, removing landline phone numbers that are non-listed and have a prescreen result of busy or fax, and capping the number of calls to a landline phone with only a no answer disposition to 10. We will examine these adjustments on the calling results to asses the impact on cost and response rate including comparisons among states. We will also discuss the forward-looking implications of our findings for conducting dual frame telephone surveys.

May 13, 2016
Concurrent Session F

Media Effects on Public Opinion

Partisan Media's Effects in Shaping Americans' Views about Pope Francis

Heather Akin, *University of Pennsylvania*
Nan Li, *Annenberg Public Policy Center, University of Pennsylvania*
Dietram A. Scheufele, *University of Wisconsin - Madison*
Kathleen Hall Jamieson, *Annenberg Public Policy Center, University of Pennsylvania*

Pope Francis's encyclical Laudato si', the Catholic document calling for a response to global warming, was the first time a preeminent religious leader took an official stance on climate change. How would the American public respond to the head of the Catholic Church advocating explicitly for environmental stewardship? Previous research suggests that audiences, and partisans in particular, avoid cognitive dissonance by seeking information affirming pre-existing views (selective exposure) or by engaging in belief-consistent information processing (motivated reasoning). We therefore hypothesize that the Pope's call for climate change action would be well received by those who were already concerned, but that partisans skeptical about climate change would be swayed by messages discrediting the Pope and his claims, potentially further polarizing the electorate. We used data from four representative samples of U.S. adults, including two cross-sectional surveys and two panels, to analyze how exposure to partisan media steered favorability toward the Pope post-encyclical. Our data indicate the encyclical's message did not change net favorability toward Pope Francis. Pope favorability declined among Republicans, increased among Independents, and remained high among Democrats—a pattern that also held in the sub-sample of Catholics. Our analysis of the informational and cognitive mechanisms that influence opinions of the Pope shows that liberal media use is related to endorsement of the Pope's statements about climate change, while conservative media use is related to disagreement. Further, we find that partisanship has an indirect effect on approval of the Pope, through preference for ideological media sources and (dis)agreement with the stances on climate change presented in the encyclical. We conclude by discussing the implications of these findings on public opinion and communication on climate change in polarized opinion environments.

How Millennials Get News: Inside the Habits of America's First Digital Generation

Jennifer Benz, *NORC at the University of Chicago*
Trevor Tompson, *NORC at the University of Chicago*
Tom Rosenstiel, *The American Press Institute*
Kevin Loker, *The American Press Institute*
Emily Alvarez, *NORC at the University of Chicago*
Dan Malato, *NORC at the University of Chicago*
Becky Reimer, *NORC at the University of Chicago*
David Sterrett, *NORC at the University of Chicago*
Jennifer Titus, *NORC at the University of Chicago*

The Millennial generation has been subject to critiques in many areas with researchers and social critics alike expressing concern that these "digital natives" are civically uninterested, passive, or even "newsless." A comprehensive study of adults ages 18 to 34 conducted by The Associated Press-NORC Center for Public Affairs Research reveals that while Millennials consume news and information in strikingly different ways than previous generations, the vast majority are far from newsless. Millennials' methods of obtaining information via social networks and on mobile devices blend news with social connection, problem solving, and entertainment. A majority of Millennials report that keeping up with the news is important to them, many get news daily, and virtually all Millennials regularly consume a mix of content, including: hard news, lifestyle news, and practical "news you can use". Contrary to the notion that Millennials believe everything on the internet should be free, results of this study show that the vast majority of these young adults regularly use paid content for entertainment or news. Additional analyses reveal that the generation is not a monolith showing important intra-generational differences based on race and ethnicity, age, and news gathering habits. The findings provide clarity for news publishers and anyone interested in reaching and engaging this generation in civic affairs. The findings are based on a nationally representative survey of 1,045 adults age 18-34, which was conducted from January 5 through February 2, 2015. In addition to the survey, the research included a qualitative component with semi-structured in-person interviews with 10 small groups of Millennials to understand their news habits, which helped provide a richer context as to how news fits into their daily lives.

Do Muslims Pose a Threat to American Values? Symbolic Threat Mediates the Link Between News Media Messages and Support for Anti-Muslim Policies

Stewart M. Coles, *University of Michigan*
Muniba Saleem, *University of Michigan*

Researchers examining anti-Muslim attitudes have largely regarded this prejudice as a product of a fear of terrorism. This is in direct contrast to research on prejudice against Blacks, which is typically associated with the perception that members of the group violate core American values. But individuals who hold anti-Muslim prejudice may also think that Muslims' beliefs are incompatible with American values even when there is no physical threat. In this respect, anti-Muslim prejudice may be two-pronged, operating on both a material and a symbolic level. The current study is designed to assess whether anti-Muslim prejudice operates on a symbolic level. To accomplish this, we conducted an online survey-experiment among both undergraduate students (N=216) and Amazon Mechanical Turk users (N=176). Respondents were randomly exposed to one of three videos, depicting Muslims in either a positive, neutral, or negative light, or a control condition with no video. They were then asked a series of questions about how American Muslim-Americans were, as well as about each respondent's support for civil restrictions targeting Muslim-Americans. In particular, we examined whether perceptions of

Muslims as American mediated the impact of the videos on support for civil restrictions. Results revealed that support for civil restrictions was highest among those in the negative and no-video conditions, followed by the neutral and positive video conditions respectively. Results further revealed that perceptions of Muslims as un-American and perceptions that Muslims were different from other religious minorities in the U.S. mediated these effects. These findings lend support to the notion that media stereotypes of Muslims can influence non-Muslim Americans' support for policies that harm Muslims in America as a result of increased symbolic threat.

From Suppressive to Proactive: The Chinese Government's Control Strategies Over Media Coverage in the Area of Popular Protests - The Evidence from Wukan Incident

Chao Zhang, *Tsinghua University*
Shaowei Chen, *Tsinghua University*

Abundant literatures have studied government's censor behavior and find that the Chinese censorship effort is still the largest in the world. However, in recent years, news media coverage on popular protest, which is thought to be the death knell of authoritarian regime, has been rising dramatically. How to make sense of the governments' seemly contradicting behavior? Several theories have been developed to address this puzzle. The dominant argument is that the Chinese government has abandoned the previous suppressive approach and adopted a more proactive one in response to the increasingly unimpeded information flow as well as the urgent need to rein in the local government officials. Nevertheless, the existing theory is quite elusive in terms of the meaning of proactive approach, i.e. what's the specific "proactive" strategies adopted by Chinese governments.This paper complements and extends previous research by two sides. Firstly, we explicitly discuss the strategies adopted by the governments. Secondly, we offer the first large-sample-based empirical evidence to test the theory mentioned above. We collected the complete news coverage on Wukan incident—an landmark popular protest incident in China and set up our original database by content analysis. Logit regression is applied in statistical analysis.We assert that even if the news media coverage cannot be suppressed, the Chinese governments still hold the ability to control the content of reports. More specifically, to maintain the social stability, the Chinese governments incline to intervene with the way news media evaluate governments' behavior and the reporting framework in popular protests. We further developed two empirical hypotheses: 1) the government-controlled party newspaper and the relatively free commercial newspaper would differ in the way they frame popular protests; 2) negative evaluation towards governments is more likely to appear in commercial newspapers. Our empirical analysis verified the aforementioned hypotheses.

Innovations with Existing Technologies

Implications of Requesting GPS Data when Recruiting Respondents for a Travel Study

Josh De La Rosa, *Abt SRBI*

Sensor data has the promise of improving the quality of survey data. Sensor data can passively collect information with minimal cognitive burden from the respondent. However, due to privacy concerns or hesitation to adapt the technology, some respondents are unwilling to participate in a study where they are required carry a device which collects data passively. As part of a household travel survey, respondents were assigned into two groups; group one was asked to log their travel using a

diary, and group two was asked to use a diary and carry a Global Position System (GPS) device. This study compares the willingness to participate in a study between the households not asked to carry GPS device, and households asked to carry a GPS devices.

The Validity of Collecting Data Using Short Message Service (SMS): Experimental Evidence from Four African Countries

Charles Q. Lau, *RTI International*
Ansie Lombaard, *TNS*
Melissa Baker, *TNS*
Joseph Eyerman, *RTI International*
Lisa Thalji, *RTI International*

Short-message service (SMS, or "text messaging") surveys offer an inexpensive and rapid new data collection mode. SMS surveys are especially attractive in developing countries, where mobile phones are widespread. However, the validity of SMS surveys is not well understood. Coverage errors may arise due to illiteracy or a lack of mobile phones. Non-response errors may emerge, particularly among lower income people – who disproportionately have poor phone signals, problems charging phones, and financial concerns about participating. Measurement errors are also a concern due to short questions and small phone screen sizes. However, we are unaware of any research that investigates these issues. Our study investigates whether SMS surveys can produce representative samples and accurate measurements. We conducted SMS surveys with 9,697 adults in Kenya, Nigeria, Ghana, and Uganda. Surveys asks questions about socioeconomic status and technology use, and are based on a stratified random sample of mobile phone numbers in each country. To investigate how survey design affects non-response errors in Kenya, we randomized the survey length (16 questions, 8 questions), incentive ($0.50, $1.50), don't know responses (included, excluded), and the day/time of invitation. Our analysis addresses three main questions. First, how representative are SMS surveys? In each country, we compare the SMS sample composition against Census data and two gold standard, in-person surveys. We also consider how survey design (survey length, incentive, don't know, invitation day/time) affect non-response errors. Second, we explore how the common practice of sharing SIM cards and having multiple SIM cards affects sampling weights and survey statistics. Third, we investigate measurement quality. We study primacy effects by analyzing "select all that apply" questions with a randomized order of response options. We also compare the levels of "don't know" responses between SMS surveys and our gold standard in-person surveys.

Implementing Bring Your Own Device in a Survey Organization

Jessica L. Holzberg, *U.S. Census Bureau*
Casey M. Eggleston, *U.S. Census Bureau*

When a survey is administered in a Computer Assisted Personal Interview (CAPI) mode, technological costs can comprise much of the overall budget, as a survey organization needs to purchase devices such as laptops, tablets, or cell phones for interviewers to use in the field. At the U.S. Census Bureau, these costs are especially pronounced. In the upcoming 2020 Census, the Census Bureau will need devices for hundreds of thousands of temporary employees to use during nonresponse follow-up (NRFU) efforts. To mitigate these costs, the Census Bureau has been conducting research into the feasibility of a Bring Your Own Device (BYOD) initiative, which would ask enumerators to use their own personal smartphones. In order to evaluate the viability of BYOD, the Census Bureau has conducted a preliminary in-house feasibility test, a field test, focus groups, and interviewer surveys. We have also

administered questions about BYOD on a nationally representative random digit dial (RDD) telephone survey. Some of this research has focused on the technical feasibility and realized cost savings of BYOD. In this paper, we examine opinion about BYOD from two perspectives: 1.   How do current and potential interviewers perceive BYOD? Are people willing to use their own devices for Census Bureau work? 2.   How does the public feel about interviewers using their own devices to collect personal information? We find that while interviewers can likely be identified and persuaded to participate in BYOD, the public's perception of the data collection is likely to be negative when aware of the existence of BYOD. This perception could potentially be lessened if respondents receive carefully crafted messaging, but this remains an area for further investigation.

Innovative Quality Assurance Use of GIS Data in Field Surveys
Marsha A. Hasson, *Westat*
Tammy Cook, *Westat*

As the ability to collect geospatial data has become more accurate and ubiquitous, researchers are turning attention to other uses of this information to support operations management and quality assurance. Data exploration and visualization tools can support improvements in efficiency, level of effort, accuracy and time to discovery for project management staff, quality assurance specialists, and analysts. This approach is based on importing GIS data from mobile devices and field laptops and combining it with other study information (e.g., visit logs, questionnaires, time and expense). A visual review of the interviewer's day superimposed on a map shows the tasks performed in location/time context, among other features. This rich data can be used to review efficiency of routes taken, confirm time and expense reports, and other quality assurance information. One example is the use of this data to identify interviewer falsification of study activities. Previous studies have explored various statistical or pattern approaches to identify at-risk interviewers based on various indices. Other common approaches include reviewing recorded interviews and re-interviewing respondents. These approaches may take days to weeks to finalize and fraudulent activity may continue during that time. The cost to detect, confirm, remediate the data, and re-field cases can be substantial. Geospatial data offers a quick, inexpensive way to detect falsified interviews across 100% of the sample. Similarly, expenses can be verified by comparison to time and distance traveled. Analysis of two major field surveys demonstrates that the use of geospatial information succeeded immediately in reducing the time to discovery and level of effort associated with falsification detection and remediation. A data exploration and visualization tool for geospatial information is a valuable asset in ensuring the quality of survey data and reducing costs associated with interviewer falsification.

Data Quality Issues for Online Surveys

Impact of 'Don't Know' Options on Attitudinal and Demographic Questions
Larry Osborn, *GfK Custom Research*
Nicole R. Buttermore, *GfK Custom Research*
Frances M. Barlas, *GfK Custom Research*
Abigail Giles, *GfK Custom Research*

In our research, we explored factors related to nonresponse in online panel surveys, in which respondents' willingness to answer sensitive questions may be affected by their long-term relationship with the panel. We initially examined the effect of providing a "don't know" response in follow-up prompts for nonresponse to questions eliciting opinions on sensitive social issues. The results indicated

that offering a no-opinion response following an initial nonresponse appears be an effective strategy to minimize nonresponse in an online environment. Our next step was to examine the application of our initial findings to the online collection of demographic data – specifically income data. Much of the prior research into item nonresponse for income questions has focused on in-person, mail, and phone surveys, which often have missing rates of 20% or even higher (Moore, Stinson & Welniack, 1999). Thus, we conducted a second study in which respondents were randomly assigned to one of three versions of a household income question to test various ways of presenting the response options to a close-ended question. We examine correlates of missing income data and report the results of the experiment, identifying the version with the lowest rate of missingness. The results demonstrate that the inclusion of a "don't know" response and how it is presented have significant implications for the rate of missingness. In addition to reporting the nonresponse rate for each of the experimental conditions, we investigate the validity of the data collected under each condition.

Assessing Changes in Coverage Bias of Web Surveys as Internet Access Increases in the United States

David Sterrett, *NORC at the University of Chicago*
Dan Malato, *NORC at the University of Chicago*
Jennifer Benz, *NORC at the University of Chicago*
Trevor Tompson, *NORC at the University of Chicago*
Ned English, *NORC at the University of Chicago*

The rising costs and declining response rates of traditional survey modes have spurred many research organizations to conduct surveys online. At the same time, less expensive broadband connections and the popularity of smartphones have made it easier for many Americans to access the internet. The General Social Survey (GSS) shows the percent of adults who use the internet rose from 69 percent in 2006 to 86 percent in 2014. The boost in access has reduced some concerns about the representativeness of internet surveys. However, there remains little research into coverage bias, which occurs if those not in the sampling frame differ from the target population on a variable of interest. Coverage bias depends on both the proportion of population included in the sample frame and the correspondence between the sample and targeted population. This raises an important question: With the increase in internet access, has there been any change in the potential coverage bias of internet surveys? In order to assess coverage bias, we analyze the GSS to determine whether those with internet access and those without it have become more or less similar in the last decade. We calculate the potential coverage bias of internet-only surveys over time for a variety of health, economic, social, and political issues, and use regression analyses to determine whether controlling for demographic factors mitigates differences between those with and those without internet access. The results illustrate that potential coverage bias has decreased for some issues (e.g. confidence in Supreme Court), increased for other topics (e.g. support for legalizing marijuana), and remains a substantial source of potential error for several issues (e.g. support for abortion), even when adjusting for demographics.

Squeaky Clean: Data Cleaning and Bias Reduction

Frances M. Barlas, *Gfk Custom Research*
Randall K. Thomas, *GfK Custom Research*
Mansour Fahimi, *GfK Custom Research*
Nicole R. Buttermore, *GfK Custom Research*

Many researchers have argued that, to improve accuracy, we should clean our data by excluding cases from analyses when participants have demonstrated sub-optimal behaviors, such as speeding, egregious non-differentiation on grid questions, or lack of item completion. While Thomas (2014) showed that eliminating small proportions of such respondents (less than 5%) did not significantly affect survey results, a question arises as to how much data cleaning is too much? For this research, we analyzed data from three different surveys that included items for which external benchmarks existed from national probability-based sample surveys. Survey 1, was the Foundations of Quality (FoQ2) study sponsored by the Advertising Research Foundation, which included 57,104 participants from 17 different non-probability sample providers. This survey included 24 national benchmarks used to investigate bias. Survey 2 included 2,563 participants from non-probability online samples and 1,297 participants from the KnowledgePanel, a probability-based online panel. This survey had 36 benchmarks to investigate bias. Survey 3, with over 10,000 participants from KnowledgePanel, also had 36 benchmarks for bias investigation. We looked at how broadening the data cleaning criteria to exclude an increasing number of participants affected bias and external validity. The results from Survey 1 indicated that while eliminating up to 5% of respondents did not affect bias (i.e., no increase or decrease), excluding a larger proportion of respondents increased bias. We are currently analyzing data from the other two surveys to assess average bias for both probability and non-probability samples based on comparable cleaning gradations. We will also explore the demographic composition and attitudes of those flagged for deletion due to the various data cleaning criteria. Findings thus far indicate that some cleaning might not affect data estimates and correlational measures, however, excessive cleaning protocols may actually increase bias, the opposite of the intended effect.

Trap Questions in Online Surveys

Laura Wronski, *SurveyMonkey*
Mingnan Liu, *SurveyMonkey*
Colleen McClain, *University of Michigan*

As more surveys move online, new concerns about data quality must be addressed. Trap questions are a common way of ensuring data quality in online market research surveys, although little academic literature exists to verify their effectiveness. Trap questions are questions inserted into surveys that are designed to "trap" respondents who are not answering accurately. Examples of trap questions include those that instruct the respondent to select a particular answer option (i.e. "Please select B"), or that ask a simple knowledge or logic question that any respondent should be able to answer. Trap questions serve a similar purpose as other data quality checks, verifying that the respondent isn't using heuristics or lying outright in order to complete the survey. We conducted three survey experiments through SurveyMonkey with several different trap question conditions. Experiment 1. We examined whether data quality changes with the addition of one or more trap questions or an introductory block of text that announces the presence of data quality checks. Experiment 2. We examined whether the difficulty of the trap question (either moderately or very difficult) and the placement of the trap question (towards the beginning or the end of the survey) affected the percent of respondents who were "trapped." Experiment 3. We examined the effectiveness of different trap

question wordings, either announcing the trap question at the beginning of the question or at the end. We also tested two captcha-like trapping questions. We compare responses to our trap questions in each condition with traditional data quality indicators, including straight-lining, speeding, numeric rounding, response consistency, and open-end response quality. Our results will provide evidence for the usefulness of trap questions as a measure of data quality, and can help guide survey designers in determining when and how to use trap questions when writing new surveys.

New Approaches to Total Survey Error

Are Interviewer Effects on Interview Length Related to Interviewer Effects on Straight-lining in the European Social Survey?
Katrijn Denies, *KU Leuven*
Caroline Vandenplas, *KU Leuven*
Koen Beullens, *KU Leuven*
Geert Loosveldt, *KU Leuven*

Many – if not all – face-to-face surveys are subject to interviewer effects on a range of outcomes. The European Social Survey (ESS) is no exception. Previous research has shown, for instance, that ESS interviewers can have a substantial impact on their respondents' tendency to straight-line (Beullens & Loosveldt, 2013) and can systematically conduct considerably longer or shorter interviews (Japec, 2005; Loosveldt & Beullens, 2013). Interviewer effects on both characteristics (length and straight-lining) can reflect a lack of a standardized approach, i.e., the interviewers have developed their own strategies rather than adhering to general guidelines. This, in turn, can reduce data quality. This paper is the first to investigate both phenomena simultaneously. Applying multilevel covariance analyses to ESS round 7 data, it reveals whether interviewers who systematically influence the interview length to a larger extent than other colleagues, are also more likely to systematically allow their respondents to keep choosing the same answer on scales. The analyses take into account country-level clustering as well as interviewer-level clustering. They furthermore control for several respondent characteristics that could be related to interview length or straight-lining, such as age, education, and mother tongue. The inclusion of these characteristics disentangles real interviewer effects from apparent interviewer effects that are, in fact, related to the area or to characteristics of the allocated group of respondents (e.g., longer interviews because the interviewer happens to be assigned to a neighborhood where most respondents do not have the interview language as their mother tongue). Interview characteristics are also taken into account, e.g., routing because more questions can obviously lengthen interviews, and interview language because some languages are more concise than others. This study emphasizes a need to acknowledge possible interviewer effects in substantive research, and it expands the knowledge-base on the relationship between interviewer effects on different characteristics.

Motivated Underreporting and Response Propensity: Do Persons Likely to Respond Give Better Answers to Filter and Eligibility Questions?
Ruben L. Bach, *Institute for Employment Research (IAB)*
Stephanie Eckman, *RTI International*
Jessica Wengrzik, *GESIS – Leibniz Institute for the Social Sciences*

Does the effort to bring in reluctant respondents decrease total survey error? Or does it increase measurement error due to low quality responses given by these reluctant respondents?

Evidence for a relationship between measurement error and nonresponse error is still mixed. To get a better understanding of this relationship, we look at an unexplored type of measurement error, motivated underreporting. Motivated underreporting refers to the tendency to give inaccurate answers to speed through an interview and may result from asking filter, looping or eligibility questions in a survey. Using five datasets conducted in different countries and in different modes, we analyze whether motivated underreporting is worse among those respondents who were the least likely to respond to the survey. To identify the reluctant respondents, we build response propensity models using boosting, a technique of statistical learning. We do not find evidence that motivated underreporting is worse among reluctant respondents in filter questions but in eligibility and looping questions.

Multi-client Household Panel Quality: The Case of AmeriSpeak

Martin Barron, *NORC at the University of Chicago*
J. Michael Dennis, *NORC at the University of Chicago*
Nada Ganesh, *NORC at the University of Chicago*

Successfully implemented, multi-client household panels can offer a compelling combination of scientific rigor, cost-effectiveness, and speed of data collection by empaneling a representative household sample. Such probability-based panels are capable of producing valid population estimates with appropriate confidence intervals. Household panels also bring with them a variety of quality challenges. The design adds a stage of recruitment (you must be recruited both to the panel and to a particular survey), thus reducing response compared to one-stage cross-sectional studies. Panelists, because they participate in multiple surveys, potentially become conditioned and be more prone to "satisficing" respondent behaviors as manifested by "straight lining," speeding, or otherwise providing low quality responses. And, of course, panel designs must deal with attrition as members exit the panel. In this paper we explore the quality of one probability based survey: NORC's AmeriSpeak. AmeriSpeak, a multi-client panel developed and funded by NORC at the University of Chicago, is now entering its second year of fielding surveys. Over the course of its first year, NORC has conducted dozens of client surveys and now has paradata on sample performance for sample recruitment and for AmeriSpeak client panel surveys and the quality of responses. Using paradata—including response rates, comparisons to external ancillary data, and calibration and weighting data--we explore the overall quality of estimates from the Amerispeak panel.

Re-interview Power-cost Tradeoffs for Identifying and Adjusting Measurement Effects in Multi-mode Surveys

Thomas Klausch, *GESIS – Leibniz Institute for the Social Sciences*
Barry Schouten, *Statistics Netherlands*
Buelens Bart, *Statistics Netherlands*
Van den Brakel Jan, *Statistics Netherlands*

A re-interview in different survey modes can be used to estimate measurement effects in multi-mode surveys. Such an identification is needed when a multi-mode design is considered beneficial for selection effects against a single mode but disadvantageous for measurement effects against a single mode. The estimated measurement effects can be used to inform data collection designers and to adjust measurement effects. In a sequential mixed-mode survey with two modes the re-interview amounts to re-approaching both respondents and nonrespondents to the first mode with the second mode. In the re-interview, a set of key questions from the mixed-mode survey is repeated. Based on the re-interview data an estimator, called inverse regression estimator, is applied to estimate the

measurements not observed under the measurement benchmark mode. The measurement benchmark mode is a question-specific choice and reflects the mode deemed most accurate. In this presentation, we first summarize earlier research on the design and the estimator. We then present new results on simulations in which we varied the size of the re-interview sample to evaluate the question, whether the re-interview design can be cost-efficiently implemented in general population surveys. It is obvious that the re-interview to respondents in the first stage causes additional costs, which depend on the number of re-interviewed respondents and the re-interview mode. However, the size of the re-interview sample determines the accuracy of adjusted estimates. Whereas it is an option to invest additional budget for the re-interview, this choice is only viable if the additional costs are small. When survey budgets are fixed, furthermore, the overall sample size of the mixed-mode survey needs to be decreased compared to a survey without re-interview. This choice causes an increase in sampling error. Given these constraints, we provide evidence supporting the choice of re-interview sample size in practice.

Evaluating New Methods in Election Polling

Using Precinct Level Election Results and Voter Registration Databases to Address Challenges In Down Ballot Election Polling
Jonathan M. Robinson, *Catalist, LLC*

The bulk of election polling is done at the statewide or national level. While much of this has to do with the lower levels of interest in those elections, the state of the art RDD survey design is generally not well suited for this kind of research. Voter registration databases on the other hand are much more amenable to this kind of research, containing detailed political geographies down to the precinct level and also in many cases a great deal of person level attributes from official, commercial, and other ancillary data sources. A common use of voter registration databases in survey research is for voter validation to improve analysis and likely voter models. This is enabled by the fact that voter registration databases include administrative data on whether a registered voter has cast a ballot in an election. However in down ballot races from House of Representatives to State Legislature, the number of ballots cast as a whole rarely equal the total number of votes cast in these races. This phenomenon, known as ballot-rolloff is a unique difficulty when it comes to measuring opinion in these kind of electoral contexts. How do we address this challenge? By matching precinct level election returns to the voter file, along with survey data from various years and election contexts we: 1. Explore how self reported roll-off correlates with actual rolloff at the precinct level 2. Identify which questions (if any at all) best capture rolloff as measured at the precinct level 3. Propose methods to improve estimating vote choice in down ballot elections using ecological regression modeling of the precinct data to correct for rolloff propensity.

Disentangling Total Error, Bias and Variance in Election Polls
David Rothschild, *Microsoft Research*
Sharad Goel, *Stanford University*

It is well known among both researchers and practitioners that election polls suffer from a variety of sampling and non-sampling errors, often collectively referred to as total survey error. However, reported margins of error typically only capture sampling variability, and in particular, generally ignore errors in constructing the sampling frame (e.g., errors due to uncertainty in who will vote). We empirically analyze 4,221 polls for 608 state-level presidential, senate, and gubernatorial races between 1998 and 2014. We find average survey error (RMSE) is approximately 3.5%,

corresponding to 95% confidence intervals of +/- 7\%, twice the width of most reported intervals. Using hierarchical Bayesian models, we decompose survey error into race-level bias and variance terms. We find average absolute race-level bias is about 1.5%, indicating that polls for a given race often share a common component of error, possibly because the polls, even when conducted by different polling organizations, rely on similar sampling frames. We further find that average race-level variance is higher than what most reported margins of error would suggest. We conclude by offering recommendations for incorporating these results into polling practice.

Calibration of Non-probability Samples to KnowledgePanel for Election Polling

Robert Benford, *GfK Custom Research*
Randall K. Thomas, *GfK Custom Research*
Mansour Fahimi, *GfK Custom Research*
Frances M. Barlas, *GfK Custom Research*
David Pace, *The Associated Press*
Emily Swanson, *The Associated Press*

Prior research collaborations between the AP and GfK have found promising results based on innovative methods for predicting the Likely Voter Turnout (LVT) and other election results in Georgia and Illinois when using a combination of survey data secured from a probability-based KnowledgePanel (KP) sample and a non-probability sample from Opt-in (OP) panels. Prior research employed a univariate calibration methodology to correct for the higher tendency of OP respondents for early adoption of new products. However, GfK finds that a more robust multivariate calibration methodology can further reduce the inherent biases associated with OP samples. This research has investigated applications of a new calibration methodology to predict off-cycle November election results in Mississippi and Kentucky using different likely voter models. The authors will present highlights of this research for calibration of opt-in samples to improve the accuracy of survey estimates for seven races in Kentucky and Mississippi. Predictive failures were common with these races, particularly for the Governor's race in Kentucky. Further, results will be presented based on a simulation study whereby smaller contributions are allowed for the KP sample component to test whether smaller probability-based sample contributions are viable in situations when larger contributions from KP are not feasible.

Impact of Incentives on Response Rates

Using Cash Incentives to Help Recruitment in a Probability Based Web Panel: The Effects on Sign Up Rates, Recruitment Cost and Sample Composition

Ulrich Krieger, *University of Manheim*

In this paper we present results from an controlled experiment on incentive use during the reminder process for recruitment into the German Internet Panel (GIP), an online panel based on a probability sample. Gaining consent and cooperation from respondents selected to be future panel members is a difficult task, dropout can occur at many stages of the survey process. To help maintain cooperation of respondents during this process, cash incentives can be used.A randomized experiment conducted during the recruitment into the German Internet Panel allows us to examine of the effects of incentive treatment on recruitment success. Following a face-to-face recruitment interview, respondents were invited to sign up online and register for the GIP via mail. If respondents failed to sign up, a reminder letter was sent. With this letter, a random group of the target respondents received a 5 Euro cash incentive. The main effect of the experiment is plain and simple, cash incentive group was

more likely to register to the panel. However, using data from the recruitment interview on sociodemographics, voting preference and income we can go beyond looking at signup rates. We can observe possible biases and differences in sample composition obtained in the online panel in respect to incentive treatment. In addition, total recruitment cost can be compared between the incentive group and the control group. With these results we can describe the effects of incentive use in a reminder process both on cost as well as quality of the resulting sample.

Understanding Incentives for Completion: Charitable Contributions, Direct Payments and Choice
Catherine Wilson, *23andMe*

A growing body of research that indicates that charitable donations often have greater value to recipients than cash rebates of an equivalent amount, suggesting that a payment to a charity on behalf of a respondent may be worth more to the respondent than the same dollar amount payment made directly to the respondent. Charitable donations may minimize some potentially negative impacts of cash payments on survey taking behavior, such as poorer data quality due to respondents rushing through surveys to complete them and collect payment, or reduced response rates due to the altruistic motivation to take part in research being replaced by monetary demand. 23andMe conducts research into genetics, ancestry, and various health conditions and diseases. The primary means by which we collect data for these projects is through participant self-report in online surveys. We are interested in finding ways to maximize data quality and survey response rates while minimizing longitudinal attrition. Interviews with our research participants have shown us that they are primarily motivated to take surveys for altruistic reasons. By reporting and updating their health histories they are contributing to research toward genetic discovery that they believe is valuable and beneficial to both themselves and others. At the start of 2015, 23andMe embarked on a unique experiment with 20,000 research participants to study the impact of offering different kinds of compensation for completion of a longitudinal study. Each of the participants in the study were randomly assigned to 1 of 6 groups, which varied in whether they were offered direct money compensation, charitable contributions of different dollar values, or a choice between each. The results of this experiment, which concludes in January 2016, will provide evidence as to best practices in compensating participants using direct payment or a contribution to a non-profit, and the monetary amount of each.

Incentive Effects on Respondent Characteristics and Survey Responses
Lindsey Witt-Swanson, *University of Nebraska-Lincoln*
Jolene D. Smyth, *University of Nebraska-Lincoln*
Kristen Olson, *University of Nebraska-Lincoln*

A considerable body of research shows that prepaid incentives increase response rates in mail surveys (e.g., Church 1993, Singer &Ye 2013). This increase can sometimes be due to an increase in a certain type of respondent that may have participated at lower levels without the incentive (Petrolia and Bhattacharjee 2009, Stinger and Kulka 2002), thus impacting nonresponse bias in survey estimates (Groves 2006). However, most incentive studies focus on response rates; only a small handful have assessed the impact of incentives on nonresponse bias (Groves and Peytcheva 2008). This paper will report the results of a prepaid cash incentive experiment ($0 vs. $1) conducted in the 2014 Nebraska Annual Social Indicators Survey (NASIS) (n=1,018). The NASIS is an annual mail, omnibus survey of Nebraska adults with sample drawn from the USPS Delivery Sequence File. Response rates by experimental treatment will be reported, but analyses will also compare the demographic make-up of the completed samples for the incentivized versus non-incentivized treatments and compare both

treatments to American Community Survey population estimates in order to understand if the incentive impacted the makeup of the completed sample in ways that would be expected to impact nonresponse bias. Primary demographics of interest include age, gender, race, education level, income, marital status, and presence of children in the home. Respondents and nonrespondents from each group will also be compared by region as a way of measuring nonresponse bias. Responses to the 2014 NASIS will also be compared by experimental group in order to understand if the incentivized group responded differently than the non-incentivized group controlling for demographics. The 2014 NASIS asked questions on a range of items including natural resources, underage drinking, safety, vaccinations, the Affordable Care Act, and plant management.

Early-bird Incentives: Results From an Experiment to Determine Response Rate and Cost Effects

Jessica De Santis, *Mathematica Policy Research*
Ryan Callahan, *Mathematica Policy Research*
Shawn Marsh, *Mathematica Policy Research*
Irma Perez-Johnson, *Mathematica Policy Research*

It is well documented that incentives can help achieve high response rates by increasing the sample members' propensity to respond. By significantly reducing reminder efforts, incentive payments can help contain costs for the researcher while sharing monetary benefits with the participants. As part of the web-based follow-up survey of dislocated workers for the Self-Employment Training Demonstration for the U.S. Department of Labor, we conducted an experiment to determine the appropriate incentive amount to offer to survey respondents. During the first four months of the follow-up survey, sample members were randomized into three groups and offered either (1) a $50 incentive for completing within four weeks and $25 thereafter; (2) a $25 incentive for completing regardless of timing; or (3) no monetary incentive. In addition to measuring the effect on response rates and timing, we set out to determine whether early-bird incentives could lower the cost per completion by encouraging faster response, thereby reducing the need for additional follow-up. In this paper, we summarize key features of the incentive experiment and our fielding efforts. We found that offering incentive payments significantly increased response rates and encouraged respondents to complete the survey sooner, resulting in a lower average cost per completed survey, particularly for the graduated $50/$25 incentive scheme. After implementing the graduated incentive payment scheme across all participants, we are seeing higher response rates, faster survey completion, and lower costs per completed survey, suggesting wider benefits for similar web surveys.

"Cheaters" vs. "Good Samaritans": Predicting Who Will Cash Prepaid-incentive Checks

Robyn Rapoport, *SSRS*
Michelle M. Doty, *The Commonwealth Fund*
Elizabeth Hamel, *The Henry J. Kaiser Family Foundation*

Monetary pre-incentives in the form of checks are frequently used in surveys of physicians and other professionals to help induce cooperation and thereby enhance response rates. Considered a best practice in conducting research among physicians, in particular, sending pre-incentives along with an invitation to participate has been associated with meaningfully higher response rates than promised incentives (e.g., Delnevo et al. 2004). A significant concern when mailing out pre-incentive checks, however, is that an unknown portion will be cashed by 'cheaters' – that is, those who cash the pre-

incentive check but do not complete the survey. The prevalence of this type of cheating can impact survey costs and potentially defeat the purpose of pre-incentivizing respondents. Previous research (e.g., Hogan, 2009) suggests that understanding who is likely to cash the check among those who complete and those who do not complete the survey is critical in determining the cost of conducting research with these populations. This paper reports findings from two studies fielded in 2015 with primary care providers (General Practitioners, Internists, and Pediatricians), where special attention was given to the problem of non-respondents cashing their incentive checks. Pre-incentive checks were sent to physicians in both studies. The incentive amount across the two studies differed ($25 vs. $50). The findings from these studies include prevalence estimates for this type of cheating, broken down by sample-based information such as site/physician specialty type, available modes of contact, and other known physician demographics that may be used to predict the likelihood that pre-paid incentives will be cashed by 'cheaters.' Conversely, we report on factors that may predict the likelihood that primary care doctors will complete the survey but will choose not to cash their prepaid incentive check.

Reaching the Hard to Reach

Planning for Production: Design of the Sampling Plan for the 2016-2017 National Survey of Children's Health
Jason Fields, *U.S. Census Bureau*
Scott Albrecht, *U.S. Census Bureau*
Keith Finlay, *U.S. Census Bureau*
Antoinette Lubich, *U.S. Census Bureau*
Tracy Mattingly, *U.S. Census Bureau*

The Maternal and Child Health Bureau (MCHB), within the Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services (HHS) has undertaken an important transition in the National Survey of Children's Health (NSCH). Working with the Census Bureau, MCHB plans to transition the NSCH to a household address-based sample from a landline telephone frame. The sampling approach requires identifying households with children and sampling a child from within those households to provide state-level representation of children with and without special health care needs. Based on survey requirements and initial results from the 2015 pretest, Census has developed a sampling plan for the 2016-2017 production NSCH that shifts to the Master Address File (MAF) with administrative inputs as an integrated sampling strategy. Four innovations will be enabled by this transition. First, we will create child flags from the Social Security Administration's (SSA) Numident and other sources and place child-present flags on the MAF prior to sampling. These flags are expected to increase the success in observing children in households with sample flags indicating children, above the 40% level observed in the pretest. Second, we will develop a model-based flag identifying 'internet-enabled/internet-restricted' households. Inputs for this model may include broadband coverage, prior web-survey respondent areas, and flags indicating the likelihood of electronic filing. All households will be invited to respond by web and we can tailor contact strategies so households flagged as 'internet-restricted' can divert to paper earlier than 'internet-enabled', where we will make additional attempts to interview by web. Additionally, we are considering both a sampling strata based on flags from SSA's Supplemental Security Income (SSI) program and a traditional poverty strata. This presentation further describes these innovations, reports initial evaluations of the administrative inputs, and describes the data collection design for the 2016-2017 production NSCH.

Using Adaptive Design Methods to Locate and Survey a Longitudinal Sample After Decades

Karen Grigorian, *NORC at the University of Chicago*
Tom Hoffer, *NORC at the University of Chicago*
Jill Connelly, *NORC at the University of Chicago*

Increasingly, survey research projects are utilizing adaptive design methods to minimize nonresponse bias and maximize limited budgets. However, most research focuses on adjusting contacting outreach and incentive offers based on real-time results during data collection. There is less information about how to use adaptive design strategies to optimize locating activities when trying to find longitudinal sample members before data collection outreach begins and this guidance is needed. Even with the best historical records, locating individuals can be an enormous challenge. The longer the interval between survey cycles, the more difficult and costly locating activities become, and the more likely the locating results will inadvertently exacerbate nonresponse bias. Bias due to uneven locating results is particularly concerning when some sample domains are easier to find than others. This paper discusses adaptive design methods used to prioritize the High School and Beyond (HS&B) sample for locating prior to data collection outreach to mitigate nonresponse bias. The HS&B is a national longitudinal study originally conducted as a part of the U.S. Department of Education's National Center for Education Statistics longitudinal studies program. In 1980, the study sampled and surveyed sophomore and senior students from a nationally-representative sample of 1,015 public and private high schools and conducted follow-up surveys in 1982, 1984, 1986, and 1992 as students progressed from high school into the early stages of their adult life. HS&B sample members are now well into middle age and continuing the study offers the opportunity to learn more about how high school and early adult experiences affect people's lives in their 50s and beyond. We describe the methods deployed in a contemporary HS&B follow-up project initiated in 2013 to be completed in 2015. The overriding goals established for these efforts were, first, to minimize nonresponse bias and, second, to maximize survey participation.

Tracing the Untraceable: A Look at Integrated and Flexible Locating Strategies in a Large-scale Study of Non-custodial Fathers

Jennifer Herard, *Mathematica Policy Research*
Rachel Sutton-Heisey, *Mathematica Policy Research*
Rebecca DiGiuseppe, *Mathematica Policy Research*
Cleo Jacobs Johnson, *Mathematica Policy Research*
Shawn Marsh, *Mathematica Policy Research*

The Parents and Children Together (PACT) study is one of the first large-scale evaluations of economically disadvantaged, non-custodial fathers. At baseline, 50% of fathers enrolled in the study had no earnings, 54% lived in an unstable housing situation, and 73% reported a criminal conviction. The challenges of following up with this highly mobile population are compounded by chronic unemployment and housing instability. Despite these challenges, it is important to obtain high response rates to reduce non-response bias and maintain validity (Kleschinsky, 2009; McKenzie, 2009). This paper examines the flexible field and in-house locating strategies used to reach response rate goals on PACT's Responsible Fatherhood 12 month follow-up survey. The PACT Evaluation is a rigorous 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. Although we anticipated this sample would be challenging to reach, we

learned that over time our respondents experienced dramatic life changes that made finding them extremely difficult. Some respondents were so "off-the-grid" (e.g. relatives and friends cannot find them) that traditional locating methods were ineffective. Many respondents were initially unlocatable, but over time we observed that respondents eventually rematerialized. We adapted our approach to consistently monitor respondents, capitalizing on their reemergence to get completes. Of completes to date, 57% required field locating. In this paper we share the integrated field and in-house locating strategies that were effective based on our ongoing review of cases including monitoring respondents in prison, maintaining relationships with contacts, and intensive reviews of exhausted cases to prioritize viable leads.

The 2015 Great British Polling Disaster - What Have We Learned?
Mario Callegaro, *Google*
Patrick Sturgis, *University of Southampton*
Michael Traugott, *University of Michigan*
Claire Durand, *Université de Montréal*
Scott Keeter, *Pew Research Center*
Natalie Jackson, *Huffington Post*

In the months and weeks leading up to the 2015 UK general election the polls told a consistent story; the Conservatives and Labour were in a dead heat in the popular vote. This led media commentators, party strategists, and the public to focus attention on the likely composition of a coalition government, rather than on the manifesto of the Conservative party who ultimately won the election with a lead of 6.6% over Labour. The expectation of a hung parliament in the final days of the campaign was so strong and widely held that the sense of shock and disbelief was palpable when the result of the exit poll was announced at 10pm on May 7th.

Questions were soon asked about the role of the polls in creating a false collective expectation about the likely outcome of the election and media sponsors publicly questioned the quality and value of the polling they had commissioned. Politicians suggested the polling inaccuracies had affected the outcome of the election, speculating that Labour might have done better had the polls been accurate. A private members bill was introduced in the House of Lords, proposing state regulation of the polling industry.

In response to these events, the British Polling Council and the Market Research Society instigated an independent Inquiry into the causes of the discrepancy between the final polls and the election result. In this panel, Patrick Sturgis will set out the key findings and recommendations of the Inquiry. Members of the panel will then respond to the Inquiry's findings and recommendations before opening to Q&A from the floor.

Questionnaire Design: Experiments on Response Options and Format

Multi-national Replication of Experiments on Acquiescence from Schuman and Presser

Henning Silber, *GESIS - Leibniz Institute for the Social Sciences*
Tobias Stark, *Utrecht University*
Annelies G. Blom, *University of Mannheim*
Jon A. Krosnick, *Stanford University*

Numerous experiments on questionnaire design were conducted more than three decades ago in the U.S. (e.g., Cantril 1940; Gallup 1941; Schuman and Presser 1981). Some important wording and format effects were particularly strong among people with less formal education (Narayan and Krosnick 1996) who have limited cognitive skills (Ceci 1991) and therefore tend to satisfice. Many of the current best-practice standards in the field are based on the findings of these studies. These standards are used by survey researchers and survey designers from many countries around the world as guidance for optimal questionnaire design. However, it is not clear that the results from the dated U.S. studies are still valid today and also applicable in other countries. This study tested whether the findings of the original experiments on acquiescence replicate in the U.S. today and whether they are of the same magnitude in 13 other countries including Canada, Denmark, France, Germany, Iceland, Japan, the Netherlands, Norway, Portugal, Spain, Sweden, Taiwan, and the United Kingdom. Four acquiescence experiments from Schuman and Presser's (1981) classic book were replicated with national probability samples of the general populations in each of the various countries (N = 1,000 each) in 2014 and 2015. For each country separately, we tested for the presence of acquiescence response bias by gauging whether responses distributions vary across question forms and whether such differences were most pronounced among respondents with the least formal education. We also subjected all experiments to a meta-analysis to test whether education moderated the strength of the response effects.

Number of Response Categories and Reliability of Measurement

Duane F. Alwin, *Pennsylvania State University*
Erin Marie Baumgartner, *Rice University*
Brett Aaron Beattie, *Penn State University*

There is a wide variety of views on the optimal number of response categories to use in the measurement of subjective content in survey research. One view is that the best approach is to construct composite scales based on responses to simple dichotomous questions. The idea is that such questions are easy for respondents understand and the use of composite scores make up for whatever lack of precision might result from such coarse categorization of response. Another view, based on information theory, is that questions that utilize a larger number of categories contain greater amounts of information and that more categories result in a higher level of measurement precision. This paper tests several hypotheses about the number of response options using reliability of measurement as the criterion for evaluating measuring precision. Using data from three General Social Survey (GSS) panel studies—2006-08-10, 2008-10-12, and 2010-12-14—we assess data for over 500 survey questions with

respect to the reliability of measurement. Our results indicate that, contrary to information theory, the most reliable measurement results when fewer response categories are used, and generally, there is a monotonic decline in levels of reliability with greater numbers of response options are employed. There is some suggestion in these data that the use of middle categories results in more measurement errors, but these trends are weak, and the overall conclusion of the research supports the use of fewer response categories rather than more.

What's It Going to Be? Yes or No?: Acquiescence Bias vs. Salience in Response
Randall K. Thomas, *GfK Custom Research*
Frances M. Barlas, *GfK Custom Research*
Nicole R. Buttermore, *GfK Custom Research*
Jolene D. Smyth, *University of Nebraska-Lincoln*

Elements in Yes-No Grids are endorsed at higher rates than when they occur in a Multiple Response Format (i.e., select all). Thomas and Barlas (AAPOR 2015) presented evidence that this may be due to increased consideration of each element. Asking for a response for each element to be considered will increase the access of less proximal memories (the salience hypothesis). Alternatively, Callegaro et al. (2015) proposed that acquiescence bias most likely explained heightened endorsement. Acquiescence bias results from socialization that encourages people to be agreeable leading to a greater tendency to endorse 'agree' with an agree-disagree response format or select 'yes' in a yes-no choice format. Our current study tested the divergent predictions of the salience and acquiescence bias hypotheses. With 1,629 KnowledgePanel (probability-based online panel) respondents, our study had a 2 X 5 factorial design – respondents were randomly assigned to a rating dimension (either a descriptive or agreement dimension) and to one of five response formats. Each participant rated four brands with which they had some familiarity along six different attributes (e.g., 'Is distinctive', 'is expensive', etc.). Those assigned the descriptive dimension were asked if the attributes described each brand, with response formats being either a multiple response format, dichotomous format ('Describes, Does not describe' or 'Yes-No'), or trichotomous format ('Describes very well; Describes somewhat; Does not describe'; or 'Yes, very well; Yes, somewhat; No, does not'). The agreement dimension had comparable response formats (e.g., 'Agree, Do not agree'; 'Yes-No', etc.). We found that trichotomous formats had the highest endorsements while the multiple response formats had the lowest endorsements regardless of responses used, supporting the increased consideration-salience hypothesis. In addition, the selection of 'Describes' or 'Yes' or 'Agree' were not significantly different, for either the dichotomous or trichotomous formats, disconfirming the acquiescence bias explanation.

Extreme Response Style: Scale Type and Left/Right Preferences
Allyson L. Holbrook, *University of Illinois at Chicago*
David Sterrett, *NORC at the University of Chicago*
Xiaoheng Wang, *University of Illinois at Chicago*

Extreme response style (ERS) is the tendency to select extreme responses in questions that use response scales regardless of their content. ERS is associated with individuals' demographic characteristics (e.g., Bachman & O'Malley, 1984), personality (e.g., Naemi, Beal, & Payne, 2009), and culture (e.g., Johnson, Kulesa, Chou, & Shavitt, 2005). However there is little evidence about how question characteristics might influence ERS (with the exception of evidence that ERS is positively associated with number of response options; e.g., Batchelor, Miao, & McDaniel, 2013)). We use data from the European Social Survey (ESS) to explore two important factors that may influence extreme

responding. The ESS includes a large number of 11-point semantic differential items with verbally labeled endpoints and numbered response options from 0 to 10 that provide a unique opportunity to study ERS. Some of these scales are unipolar (e.g., satisfaction) and others are bipolar (e.g., evaluation). All response scales are visually presented using showcards. We use a variety of analysis strategies including factor analysis and multilevel modeling to analyze these data to explore whether ERS differs for unipolar and bipolar response scales and to examine ERS separately for each end of the visually presented scales. We examine whether the level of ERS varies as a function of type of scale (unipolar v. bipolar) and scale endpoint (left endpoint v. right endpoint) and also explore whether the predictors and correlates of ERS vary as a function of these factors. This study will offer insights into whether ERS varies across question types and what factors are associated with different types of ERS.

Election Issues for 2016

Is This Any Way to Elect a President?
Gary Langer, *Langer Research Associates*
Gregory Holyk, *Langer Research Associates*
Chad Kiewiet de Jonge, *Langer Research Associates*
Julie Phelan, *Langer Research Associates*
Margaret Tyson, *Langer Research Associates*

With a system designed for early winnowing, major party nominees for president may be known by the time of the 2016 AAPOR conference. With a substantial share of the electorate inclined to support unconventional candidates, maybe not. Regardless, deep data will be on hand to analyze primary voters' preferences – and to discern the contours of the general election contest ahead. Our paper will synthesize our findings across the primary season, drawing upon our coverage of the election for ABC News – through a series of ABC News/Washington Post pre-election polls and our election-night analysis of the National Election Pool exit or entrance polls in Iowa, New Hampshire, Nevada, South Carolina, Super Tuesday and beyond. Our analysis will seek to elevate the discussion beyond the horse race to the key dynamics of the election. These include who participated in the primaries and caucuses, what issues and candidate attributes informed their vote choices and – drawing on our rich resource of exit poll trends back 40 years – how these factors compare with past nominating seasons. Similarly, we'll use our early general election polling to assess the race ahead, looking at voter demographics and attitudes alike to tell the still-unfolding story of the 2016 presidential election.

The Effects of Poll Proliferation on Polling Quality and Accuracy in the Aggregate

Natalie Jackson, *Huffington Post/Pollster*
Ariel Edwards-Levy, *Huffington Post/Pollster*
Janie Velencia, *Huffington Post/Pollster*

There are more pre-election polls than ever. But there are also more questions than ever about polling accuracy, quality, and methodology, meaning that more data might not necessarily be better. We use the 2016 primary election polls from our HuffPost Pollster database to investigate how polling in the aggregate has been affected by poll proliferation. As of mid-November 2015, the number of national primary polls conducted in the 2016 cycle is 25 percent higher than the number of national polls at the same point in 2012 cycle, and 60 percent higher than in mid-November of 2007. In Iowa and New Hampshire, the number of 2016 polls is on track to easily surpass 2008 and 2012 totals with two and a half months left until the first caucus and primary. Our analysis will compare polls from the last two U.S. presidential elections and attempt to explain why there has been a rise in polls conducted, who these new pollsters are, if the poll modes have changed and how these polls are affecting the aggregate estimate of our model. We will also provide an assessment of how accurate the polls are in New Hampshire, Iowa, and other early states with a substantial number of polls. We will conclude by discussing the implications of these changes and what it means for this cycle and the future of polling.

The End of the Interviewer?: Differences in Data Collection Methodologies in State Pre-election Polls

Andrew E. Smith, *University of New Hampshire*
Jennifer Dineen, *University of Connecticut*
Zachary Azrm, *University of New Hampshire*

Recent elections have seen the rise and growing importance of poll aggregators, such as pollster.com (now part of the Huffington Post), RealClearPolitics, and especially, Nate Silver and his 538 blog in framing elections. The success of 538 and others in predicting the state by state outcomes of the 2008 and 2012 presidential elections solidified their importance in media analysis and among pundits. But the 2014 mid-term elections highlighted a major criticism of aggregator sites – that their predictions are only as good as the data that go into their formulae. Most importantly, in many states, poll averages are driven increasingly by either low cost IVR or online polls and/or polls conducted by partisan political polling groups. As the number of polls conducted using alternate methodologies increases, few attempts have been made to examine their impact on polling averages. This paper analyzes the content of poll aggregation sites during the final months of presidential, gubernatorial and senate campaigns since 2008. We categorize polls by methodology, partisanship, and accuracy over to test the theory that poll aggregator averages at the state level are being increasingly driven by polls that do not meet AAPOR or CASRO best practices. Preliminary analysis shows an increase in in the percentage of IVR and on-line polls, especially in smaller states that are not typically on the radar screens of major media outlets and that IVR polls exhibit lower percentages of undecided voters and lower variances than do traditional polls.

Voter Evaluations of an All Vote-by-Mail Election

William P. Marble, *Stanford University*
Melissa R. Michelson, *Menlo College*

Vote-by-mail is an increasingly popular method of administering elections, but few studies have examined how the public perceives all-mail voting. Stein & Vonnahme find that voters approve of voting methods that increase convenience (Social Science Quarterly 2012), while Southwell and Burchett (American Politics Quarterly 2000) and Southwell (PS: Political Science & Politics 2004) report high levels of voter satisfaction with mail-in voting. California, which is considering holding all of its elections by mail, commissioned a pilot program in San Mateo County in November 2015. Voters received ballots in the mail and were encouraged to mail them back; additional voting options included 22 drop-off ballot boxes available during the month prior to Election Day and 32 Election Day Universal Polling Places (UPPs). We conducted a combination exit poll and post-election telephone survey to measure voter attitudes about the shift to an all-mailed ballot election, including perceptions of vote integrity, trust in the U.S. Postal Service, and whether the new format might advantage a particular political party. Overall, we collected 1,000 completed surveys, half via exit polls and half via post-election phone calls. In our analysis, we give particular attention to how attitudes vary between individuals who chose to mail their ballots back to the Elections Office, those who chose to vote at UPPs, those who chose to drop off their completed ballots, and those who did not vote. We examine support for holding all California elections by mail and how those attitudes differ by age, partisanship, education, ethnorace, gender, neighborhood socioeconomic status, and language (English, Spanish, and Chinese). In addition, we use the voter file to test whether turnout in the all-mailed pilot varied by those same subgroups.

Improving Congressional District Data in the Cell Phone Age

David Ciemnecki, *Abt SRBI*
Dean Williams, *Abt SRBI*
Seth Brohinsky, *Abt SRBI*

Pollsters' interest in congressional district-level information is increasingly important to help predict election outcomes. Sample vendors are able to provide reliable congressional district information for the landline frame because landline phones are tied to a geographic area. Cell phones are much more problematic. With mobile phones becoming the prevailing mode of contact, congressional district-level information is largely unavailable or inaccurate due to cell phone mobility. Many pollsters attempt to recode congressional district by asking respondents their zip code. However, nearly one quarter (24.3%) of all 41,339 US Zip Codes contain multiple (up to 5) congressional districts. As a result, it is impossible to accurately recode congressional district based on zip code for any respondent living in a multiple congressional district zip code. Complicating matters, over half of the US population (52.7%) lives in these multiple congressional district zips, meaning they're just as likely to be contacted in an RDD survey as those living in single congressional district zips. This paper, presents an approach to more accurately determine a respondent's congressional district and an analysis of its effectiveness. Using data from a large (n=20,000+) national study recently conducted by Abt SRBI, we will attempt to more accurately recode congressional district using a combination of zip code and respondent-reported congressman. We will then analyze whether this approach is reliable. Additionally, this analysis will focus on differences between respondents who live in single-congressional district zip codes and those who live in zip codes with multiple congressional districts, be able to identify any cases where the reported information is different from the provided information in the sample file, and determine if there are any predictors that can be used to identify accuracy (or inaccuracy) of congressional district sample information and predict election outcomes.

Methodological Brief - Panel Recruitment & Retention

Web Probing for Question Evaluation: The Effects of Probe Placement

Stephanie Fowler, *National Cancer Institute*
Gordon Willis, *National Cancer Institute*
Richard P. Moser, *National Cancer Institute*
Reanne L.M. Townsend, *Westat*
Aaron Maitland, *Westat*
Hanyu Sun, *Westat*
David Berrigan, *National Cancer Institute*

Though the majority of cognitive interviewing is conducted in a laboratory environment by specialized interviewers, an alternative procedure, embedded probing, was developed decades ago (Converse & Presser, 1986; Schuman, 1966). Embedded probing involves the evaluation of select items through the administration of cognitive probes within a field survey environment, relying on much larger samples than are cost-effective with standard cognitive interviewing. Embedded probing has recently resurfaced in the form of web probing, mainly prompted by the increasing availability of internet-based platforms that provide fast, inexpensive access to respondents (Behr, Braun, Kaczmirek, & Bandilla, 2013; Murphy, Keating, and Edgar, 2013). However, little guidance exists concerning specific practices. In particular, as for standard cognitive interviewing, it is unclear whether probes should be administered concurrently with the target item, or retrospectively at the end of the interview (i.e. debriefing). This study examined web-based probing in the context of questionnaire pretesting and evaluation, relying on respondents from Amazon's Mechanical Turk (mTurk), an online labor crowdsourcing platform. A questionnaire segment on walking behavior and attitudes, selected from the National Health Interview Survey (NHIS), was presented to 1,444 mTurk respondents. Probes were embedded for 4 targeted items. For example, respondents were asked: "When answering the question -- Where you live, are there roads, sidewalks, paths, or trails where you can walk? – Please say more about what you were thinking about…", then typed their response to the probe into an open text box. As an experimental manipulation, approximately half of respondents answered the probes concurrently, and the remainder responded retrospectively. We will present the results of an analysis that compares the responses obtained as a function of probe placement, by assessing quantitative metrics such as number of words entered, as well as qualitative measures, including thematic coding and rater judgment of usefulness of the information obtained.

Do Running Tally Questions Work with Smartphone Respondents?

Tom Wells, *Nielsen*

In online surveys, running tallies or constant sums are used with number boxes to help respondents sum amounts to exactly 100%, 24 hours, etc. The cumulative total is displayed on screen and updated as numbers are successively entered. In this paper, we evaluate the performance of running tally questions with smartphone respondents. We conducted a survey with a sample of online panelists from GfK's KnowledgePanel. The survey was conducted between September 17th and October 8th, 2015 and completed by 3460 respondents. 63% of respondents completed the survey on a laptop or desktop computer, 24% on a smartphone, and 12% on a tablet The running tally question asked for estimates of mobile device sharing – specifically, what percentage of the time each household member uses a specific device. Our paper focuses on smartphone respondents, since the running tally item was

most challenging, in terms of visual display, navigation, and user experience, on a smartphone. To evaluate the performance of the running tally question, we analyze differences between smartphone respondents and computer respondents in terms of item non-response, response latency, and comparability of survey responses. Our results reveal significantly longer response latency among smartphone respondents and higher item non-response, although the latter difference is not significantly significant. We also estimated a linear regression model to analyze the effect of using a smartphone on responses to the running tally question, controlling for several respondent and household characteristics. Our regression results show that use of a smartphone rather than a computer does not significantly affect estimates of mobile device sharing. These results are encouraging. Although the running tally question was not optimized for smartphones, survey estimates were not affected by the use of a smartphone. This is consistent with the general finding that there are almost no effects of device type on survey responses.

Recruiting Hard-to-Reach Populations: The Utility of Facebook and the Promise of Big Data

Orin T. Puniello, *Ketchum Global Research & Analytics*
Marc D. Weiner, *Rutgers University*
Paul C. Siracusa, *Rutgers University*
Jocelyn Elise Crowley, *Rutgers University*

Social networking facilitates new ways of contacting hard-to-reach populations. Facebook—ground zero for Big Data—provides previously unimaginable cost-effective approaches to participant recruitment. For example, in Survey Practice (2015), Schneider, Burke-Garcia, and Thomas reported successfully using Facebook for respondent tracing of former foster care youth who had been randomly selected for a longitudinal evaluation of a prior intervention. For qualitative research investigating the policy implications of "gray divorce" (divorce at or over fifty years old), we were challenged to recruit a pool of qualified respondents from which to secure 80 interview subjects of geographic diversity split evenly by sex. Although gray divorce is increasingly common—approximately one in four of all divorces now occurs among this age group—the incidence rate of individuals within the general population meeting these criteria still renders traditional recruitment methods cost-prohibitive. With limited resources coupled with the need for micro-targeting, we turned to Facebook, where, within 13 days over four waves, we recruited 178 qualified respondents from which we successfully selected 80 interviewees while preserving geographic diversity. We specified the characteristics of Facebook users who would be exposed to our solicitation: geography (United States); age group (55-65+); gender; placement of the advertisement on a mobile device (news feed); and placement on a desktop (news feed or right column). Once the respondent clicked on the Facebook ad, they were redirected to a Qualtrics screener that we administered directly. Overall, we generated 178 qualified respondents at a mean cost of $1.18 per respondent, i.e., a total recruitment cost of $210.04. Compared to other approaches, the cost-efficiency is overwhelming. This methodological brief details our interactions with Facebook, the content of our recruitment advertisements, and our decision-process in selecting the relevant targeting criteria. Ultimately, we present a cost efficient methodology for using Facebook as Big Data technology to recruit a low-incidence, hard-to-reach populations.

A Mobile Device Panel in the United States: Lessons from Collecting Data at Scale

Abhinav Agrawal, *SurveyMonkey*
Noble Kuriakose, *SurveyMonkey*

We report on our methods to construct a large scale non-probability panel to learn about mobile behavior in the United States. We provide an overview of the technical challenges to extracting usage patterns directly from phones after gaining respondents' consent to access their data. We share technological tools and the workflow we've successfully used to deal with the tremendous scale of over 500 million records per day offered by machine generated big data. We suggest a path forward with adjustment techniques when acquisition sources differ on both the demographic makeup of panelists and the exhibited patterns of mobile device usage. We compare techniques to generalize to the U.S. mobile phone population from the non-probability dataset, including raking, and share the advantages and biases presented by unweighted data. Finally, we discuss the unique challenges of adjusting and accounting for a long term panel with differential churn.

Survey Mode and Mail Method: A Practical Experiment in Survey Fielding for a Multi-round Survey

Brianna D. Sullivan, *Mathematica Policy Research*
Nancy Duda, *Mathematica Policy Research*
Karen Bogen, *Mathematica Policy Research*
Nancy A. Clusen, *Mathematica Policy Research*
Breanna Wakar, *Mathematica Policy Research*
Hanzhi Zhou, *Mathematica Policy Research*

Researchers can choose from several survey mode and contact method options when designing survey fielding plans. Each of these options is associated with known costs, yet the expected success of each fielding option, which often depends on the circumstances and preferences of a specific population of respondents, is much more ambiguous. Surveys with multiple rounds present an opportunity for researchers to identify the most successful and cost-efficient survey fielding plan in one round, to guide survey fielding in subsequent rounds. During the fielding of the 2015 Comprehensive Primary Care (CPC) Practice Survey, sponsored by the Centers for Medicare & Medicaid Services, we implemented a two-by-two survey experiment with a subset of our sample of practice managers for primary care practice sites for whom we did not have valid email addresses, only practice mailing addresses. We randomly assigned 269 practice managers to receive a letter that directed the practice to either (1) complete the survey online with a personalized web login, or (2) complete an enclosed hard copy questionnaire. We also randomly assigned the practice managers in each of those groups to be sent all outreach material using either (1) regular USPS mail or (2) FedEx next-day shipping. Though practice managers in the online and FedEx groups responded to the survey more quickly initially, over time the difference in cumulative response rates between groups disappeared. In further analyses of this experiment, we will examine the impact of survey mode (online versus hard copy) and mail method (USPS versus FedEx) on data quality, representativeness, and cost. We will recommend the best mode and procedures for future rounds of the CPC Practice Survey, as well as reflect on the general lessons learned on conducting fielding experiments in multi-round surveys.

Does Paying Out Pay Off? Short-term Panel Bonuses for All, for Some or for None?
Marla Cralley, *Nielsen*
Adam Gluck, *Nielsen*
Arianne Buckley, *Nielsen*

Incentives are often offered in research studies to help improve cooperation. Researchers face many decisions about using incentives: Should they be used at all, when should they be used, should they be contingent on performance, and what form should they take? Researchers should also balance the benefits incentives may provide with their cost. Prior research shows that contingent incentives increase participation for surveys in general (Singer et al, 1999) and for panels (McGonagle et al, 2011). In Fall 2015, we conducted a test to examine the effectiveness of a short-term promised incentive offered to panel members within participating households. The goal was to improve cooperation among demographic groups that were performing at below average rates. The test was designed to answer two research questions. First, are short-term, promised incentive bonuses an effective method for improving compliance with a specific behavior? Second, is it more effective to offer this type of incentive only to the specific low performing demographic groups in the household or to all members of the household? For this study, households with poor-complying panelists in select demographics were divided into three groups: 1) a test group where every member of the household (the targeted panelist AND others in the household) were offered the bonus, 2) a test group where only the targeted panelists were offered the bonus, and 3) a control group where no one in the household was offered the bonus. The analysis compared compliance rates between the three groups in order to develop cost-effective strategies for using incentives to improve compliance. Preliminary analysis shows that short-term, promised incentives are effective for improving compliance behavior in panels. Additionally, offering incentives only to non-cooperating members of a household appears to be the more cost-effective strategy.

Can CAPI Interviewers Accurately Predict Actual Panel Participation? Evidence from the Recruitment Process for the GESIS Panel
Bella Struminskaya, *GESIS - Leibniz Institute for the Social Sciences*
Ines Schaurer, *GESIS - Leibniz Institute for the Social Sciences*
Michael Bosnjak, *GESIS - Leibniz Institute for the Social Sciences*

Interviewer observations on the likelihood of survey participation can potentially be used for postsurvey adjustments or adaptive designs during fieldwork. However, the usefulness and validity of such observations has been questioned in past studies. The limited predictability of interviewers' observations about survey participation might reflect the fact that interviewers have few cues to base their judgement on. During a recruitment interview for a panel, interviewers have much more information to form the prediction. We study whether interviewers can accurately predict subsequent panel participation after a recruitment interview and which factors affect their predictions. We use data from the GESIS Panel, a probability-based mixed-mode general population longitudinal study, which was recruited face-to-face in 2013 (N=7599). For 6210 respondents who were willing to join the self-administered (online and mail) panel, interviewers had to rate the likelihood of panel participation. The information on actual participation in the first self-administered wave serves therefore as a "true" value to which interviewer predictions can be compared. We find a strong interviewer effect on recruitment and a weaker effect on actual participation. There is a significant effect of the interviewers' rating of the likelihood of participation on actual participation. However, the predictive validity of interviewer ratings depends on other characteristics, such as survey mode (more accurate predictions for mail respondents

than for online respondents), interviewer characteristics, interview procedure and respondent characteristics. This paper will be useful for guiding decisions about targeted interventions during fieldwork for cross-sectional studies and for panel recruitment studies.

Does Changing Monetary Incentive Schemes in Panel Studies Affect Cooperation? A Quasi-experiment on Switching from Prepaid to Promised Incentives on Participation in a Probability-based Mixed-mode Panel

Ines Schaurer, *GESIS - Leibniz Institute for the Social Sciences*
Michael Bosnjak, *GESIS - Leibniz Institute of the Social Sciences*

Monetary incentives are a universal measure aimed at fostering cooperation in surveys. Meta-analytic research based on cross-sectional surveys has shown that prepaid monetary incentives are more effective than postpaid incentives. In the context of longitudinal surveys, research on the long-term impact of the timing of incentives on respondents cooperation is scarce. As a result, incentive-related decisions in longitudinal surveys can hardly be justified in view of the available empirical evidence. When building longitudinal research infrastructures, the incentive scheme issue becomes even more prominent if financial, legal, or administrative constraints require a change in the timing of incentives, for instance from (seemingly) costly prepaid monetary incentives to promising an incentive contingent on participation. Accordingly, the overall aim of the study presented was to investigate if, and to what extent, a change in incentives schemes from a prepaid to a postpaid model affects cooperation. We focus on the initial phases of building the GESIS Panel, a self-administered mixed-mode (Web and mail surveys) access panel infrastructure aimed at hosting various longitudinal social science studies (N= 4961). Specifically, we will report findings of a quasi-experiment conducted during the recruitment process of the GESIS Panel, suggesting that the magnitudes of effects of incentive scheme switches from prepaid to postpaid on cooperation are mode specific: The likelihood of participation in the offline mode (mail surveys) is substantially more affected by an incentive scheme switch (prepaid -> postpaid) than in the online mode (Web surveys).

Experiments with Mobile Data Collection

Embedding Survey Questions within Non-research Mobile Apps: A Method for Collecting High-quality Data from Smartphone Users?

Vinay Bapna, *Unomer*
Christopher Antoun, *U.S. Census Bureau*

The rise of smartphone applications has provided survey researchers with new ways to enhance measurement. But app-based approaches to data collection typically require users to be willing to download and use a dedicated research app, which has consequences for nonresponse. The goal of this research is to describe and evaluate a technique that does not require downloading a special app. We present a strategy of embedding survey questions directly within non-research apps. The questions appear on a user's screen (in the place where advertisements are typically displayed) while they use an app. We report a large-scale evaluation of an in-app survey in a gaming app. We examine the speed of responses (e.g., number of responses per day), participation rates, demographic characteristics of respondents, and data quality. Our measures of data quality include indicators of speeding (based on item timings) and rounded numerical responses to behavioral frequency questions. We will also examine the effects of different amounts of an in-app incentive (e.g., gold coins) that provide benefits in the game. The findings can help inform strategies to improve this emerging method.

Hesitation in Socially Desirable Responses in a Mobile Phone Survey

Stefanie Fail, *New School for Social Research*
Michael F. Schober, *New School for Social Research*
Frederick G. Conrad, *University of Michigan*

When survey respondents answer sensitive questions, they can sometimes present themselves in a more favorable light than is accurate. The current study examines how respondents produce "socially desirable" answers, testing three hypotheses about response latencies as potential indicators of socially desirable responding. The study compares response latencies of answers to sensitive and non-sensitive questions in a corpus of 319 audiorecorded mobile telephone interviews from Schober et al. (2015), in which participants answered 32 questions (some sensitive and some not) on their iPhones. Half the respondents (160) were interviewed by professional interviewers and the other half (159) by an automated spoken dialog interviewing system (speech-IVR). The main comparisons are (a) within-subject comparison of response latencies by the same person to sensitive and non-sensitive questions; (b) between-subject comparisons of response latencies for selecting more and less socially desirable answers to the same sensitive questions; and (c) within-question comparisons of response latencies when sensitive questions are asked by an automated system or human interviewer. Consistent with evidence in other survey modes, respondents produced more socially desirable responses with a human than an automated interviewer. Analyses also demonstrate that between people response latencies can be significantly greater for certain options on sensitive questions; for example, respondents took longer to answer that they had sex in the relatively frequent category "2 or 3 times a week" than for less frequent categories. These findings test the generalizability of prior claims (Holtgraves 2004) that people take longer to report their traits and behaviors when there is more pressure to give a socially desirable response; additional analyses based on ratings of the social desirability of different response options allow focused examination of how and perhaps why response latencies correlate with more and less socially desirable responding.

Why Do Mobile Interviews Take Longer? A Behavior Coding Perspective

Jerry Timbrook, *University of Nebraska-Lincoln*
Kristen Olson, *University of Nebraska-Lincoln*
Jolene D. Smyth, *University of Nebraska-Lincoln*

One consistent puzzle about differences between interviews conducted on cell phones versus landline telephones is that mobile interviews take longer to complete than landline interviews (Kuusela and Notkola, 1999; Keeter and Kennedy, 2006; Brick et al., 2007; Vicente, Reis, and Santos, 2009; Lynn and Kaminska, 2012). Common hypotheses for the differences in length between these two devices are that mobile respondents are more distracted, that differences in within household selection and post-survey collection of information for incentives affect duration, and that poor audio quality on mobile devices increases duration. These hypotheses have not been fully evaluated. Yet understanding drivers of length for each device is important as interview length is a significant cost component of telephone surveys. In this paper we use response times and behavior codes from the US/Japan Newspaper Opinion Poll (n = 438) conducted during November of 2013 to investigate speech behaviors and question-answering behaviors that may explain the difference in interview length between these two devices. Preliminary findings indicate that the difference in interview length exists even after accounting for within-household selection in the landline survey and post-survey collection of information for incentives in the mobile interviews. Interview duration is 1.04 minutes longer for mobile respondents

than landline respondents (p < .001). Mobile respondents exhibit significantly more disfluencies (p = .02), make significantly more comments about poor call quality (p < .001), make significantly more inaudible utterances (p < .001), and speak at a significantly slower rate than landline participants (p < .001). We will investigate these differences further by using multivariate analyses to test the independent effect of these behaviors by device on interview duration after controlling for socio-demographic differences. This study concludes with implications for interviewing in dual-frame telephone surveys.

The Differential Effect of Mobile-friendly Surveys on Data Quality
Rachel Horwitz, *U.S. Census Bureau*

With mobile device penetration increasing across the world, so is the number of people using these devices to respond to web surveys. Research has shown that data quality is lower when people use mobile devices to respond to surveys that are not designed for their use. Therefore, many administrators have optimized their web surveys for mobile devices, or at least made them mobile friendly, so the questionnaires display to utilize all available screen space. While this change should make it easier to read the questions and select responses, the research thus far suggests that while optimization helps, it may not entirely eliminate device effects. The Census Bureau's National Content Test (NCT) and the American Community Survey (ACS) developed a mobile-friendly version of their web surveys in 2015 and 2016, respectively. This research aims to compare data quality indicators for the NCT and ACS, including breakoff rates, completion times, answer changes, and the percent of mobile device users that switch to a computer, before and after the adoption of a mobile-friendly instrument. In addition to assessing the impact of a mobile-friendly instrument on each of these surveys separately, there is great value in comparing the differential gains across the surveys. Both are national probability samples, but the survey length and content varies greatly. Specifically, the NCT takes approximately 10 minutes to complete, compared to 40 minutes for the ACS. Additionally, the ACS contains more complex questions and question formats than the NCT does. The surveys were designed and optimized in the same way, allowing for a direct comparison to inform whether developing mobile-friendly instruments can improve data quality across surveys or if the effectiveness can depend on the characteristics of each survey.

Using Paradata to Improve Data Quality

Using Early-wave Interviewer Observations to Predict Respondent Burden and Survey Nonresponse in Later Panel Survey Waves
Scott Fricker, *U.S. Bureau of Labor Statistics*
Arcenis Rojas, *U.S. Bureau of Labor Statistics*
Lucilla Tan, *U.S. Bureau of Labor Statistics*

Panel surveys are faced with the dual challenge of minimizing attrition (to maintain estimate precision and reduce the risk of nonresponse bias) and ensuring that those who do respond do so in an effortful manner. Participation decisions and response data quality both are impacted by panel respondents' perceptions of survey burden. Past research on interviewer observations about contacted sample units' attitudes and reactions to the survey request ("doorstep concerns") have found them to be predictive of final survey outcome and reported burden in the Consumer Expenditure Interview (CE) Survey, a national household panel survey. The current study extends this research to examine the extent to which doorstep concerns observed at the start of the survey panel are predictive of

respondents' doorstep concerns, survey participation, and perceptions of burden in the final CE interview. We use data-mining techniques to explore these associations, comparing model performance across different segments of the CE sample, and multiple survey outcomes. We discuss implications of our findings for further development and use of interviewer-observation paradata and measures of respondent burden, as well as possible applications of these data in a responsive design framework.

Managing Efficiency in Telephone Surveys: Insights from Survey Paradata Trends 2010-2015

Rachel Martonik, *Abt SRBI*
Tara Merry, *Abt SRBI*
Nicole Lee, *Abt SRBI*
Stephen Immerwahr, *NYC Department of Health and Mental Hygiene*
Michael Sanderson, *NYC Department of Health and Mental Hygiene*

The ever-evolving challenges associated with collecting telephone survey data are well known. People are using telecommunications technology differently, including the devices they own and how and when they use them. Landline coverage is diminishing while new Telephone Consumer Protection Act rules restrict how we can reach respondents by cell phone. And response continues to decline. Collectively, these pose a critical threat to the efficiency and accuracy of telephone surveys. In this paper, we examine indicators of these changes in production paradata over time to understand which factors most heavily influence efficiency and data quality. We conduct an in-depth analysis of production paradata from the 2010-2015 Community Health Survey (CHS), a large, dual frame telephone survey conducted annually by the New York City Department of Health and Mental Hygiene. We examine changes in frame characteristics, response patterns from call history data, and dispositions over time as they relate to indicators of efficiency and data quality. The results will be used to assess the nature of these changes and provide insight on ways to maintain efficiency in telephone survey production.

Use of Paradata to Investigate Potential Data Quality Issues in an Online Survey

Tzu-Jou Wan, *American Institutes for Research*
HarmoniJoie Noel, *American Institutes for Research*
Lee Hargraves, *American Institutes for Research*

Past research suggests that paradata can be a useful tool to detect potential sources of error in surveys. This study explores the use of online survey paradata and how it may be an indicator of data quality issues. The paradata comes from the Qualified Health Plan Enroll Experience Survey, a 107-question online survey sponsored by the Centers for Medicaid and Medicare Services. The survey was administered in English to those who provided an email address when applying for health insurance through the Health Insurance Marketplace. Using survey paradata, we identified three groups of respondents: people who broke-off part way through (break-offs), people who finished the survey but very quickly (fast respondents), and people who finished the survey but took longer (regular respondents). In this presentation, we examine the following research questions: 1) do fast respondents look more like respondents who broke-off or regular respondents in terms of time spent on each question, number of questions answered and item-level missing data?; and 2) are there any differences in personal characteristics among the three types of respondents? Preliminary analysis shows that about 10% of respondents were responding fast, defined by those who spent less than 8.5 minutes on finishing the questionnaire, whereas the average completion time was 18 minutes. The mean number of

questions answered by fast respondents is 58, compared to 74 for regular respondents and 26 in break-offs. Item-level time suggests fast respondents spent less time on long questions than break-offs and regular respondents. Race, gender and age are also associated with break-offs and speeding. The findings show how paradata can be used to identify data quality issues.

Measuring Political, Science and Religious Knowledge with Online Panel Surveys
Brian Kennedy, *Pew Research Center*
Cary Funk, *Pew Research Center*

Surveys are often used to measure factual understanding of certain topics, such as science, politics and religion, in the general public. These surveys have often tested verbal knowledge of facts, such as identifying what a light year measures, the number of women on the Supreme Court, or the names of the first four books of the New Testament. But with surveys increasingly moving to the online environment, we know little about the consequences of measuring knowledge with this mode. This project looks at possible consequences of measuring domain specific knowledge on online panels. First, it compares the measurement properties of knowledge scales by mode. It shows that on average respondents perform better online than by phone. Why? We explore one possible factor—some respondents treating questions of knowledge online as an "open-book quiz". Second, we examine the consequences of including knowledge questions with visual elements, showing that the measurement properties are similar for knowledge items with and without visual elements and finding that differences between subgroups on visual items are similar to differences on nonvisual items. Finally, we show that responses to knowledge questions online and by phone are tied to similar expected attitudes and behaviors.

New Math for Nonprobability Samples

Next Steps Towards a New Math for Non-probability Sample Surveys
Mansour Fahimi, *GfK Custom Research*
Frances M. Barlas, *GfK Custom Research*
Randall K. Thomas, *GfK Custom Research*
Nicole R. Buttermore, *GfK Custom Research*

Traditional methods of survey research that rely on Neyman's probability-based sampling paradigm are based on a number of fundamental assumptions that are becoming exceedingly unattainable. On the one hand, common methods of sampling are subject to frame coverage issues that may not be sufficiently ameliorated through post-survey weighting adjustments. On the other, response rates continue to deteriorate for all surveys, even when resource-intensive refusal conversion strategies are employed. Add in the growing pressure for cost containments and it is no wonder why alternative sampling methods are gaining acceptance. Popular among such alternatives are non-probability samples secured from online sources, however, unbiased statistical inferences cannot be generated from such samples due to their indeterminable selection probabilities and unknown coverage properties. Building on a two-year research initiative, the authors will discuss a robust calibration methodology that can reduce the inherent biases associated with non-probability samples. Results are based on national and state-level parallel surveys that employ both probability-based and nonprobability samples with identical instruments. The efficacy of our proposed methodology is assessed in terms of improved external validity of the resulting survey estimates vis-à-vis those obtained from government surveys, including the CPS, ACS, BRFSS, NSDUH, and NHIS. The bias reduction capacity of this calibration

methodology, which outperforms what can be achieved via traditional geodemographic weighting adjustments, is also demonstrated for probability-based samples from incomplete frames or those subject to high rates of nonresponse.

Quota Controls in Survey Research: A Test of Accuracy and Inter-source Reliability in Online Samples

Steven H. Gittelman, *MKTG, INC.*
Randall K. Thomas, *GfK Custom Research*
Paul J. Lavrakas, *Independent Consultant*
Victor Lange, *Consultant*

Accuracy in survey research relies on accurate measurement of key topics with a sample that represents the population of interest. A common method used to address the non-representativeness of a non-probability sample (NPS) is to implement demographic quotas to make the sample demographically resemble the intended population. However, researchers have demonstrated that even if demographics of a NPS are equated to the intended population, a NPS is neither behaviorally nor attitudinally the same. In 2013, a large scale study was conducted among 17 sample providers contributing NPS to an online survey that ran in parallel to a dual-frame phone survey. The main purpose of the research was to determine the extent of risk that existed with the use of NPS in online research. The current authors examined two approaches to improve survey quality in NP samples, including implementing demo sample selection quota controls; and using model-based selection methods, which use additional attitudinal and behavioral measures. Further, quality in terms of the proximity of results to the various U.S. benchmarks was also measured. Generally it was found that selecting samples with higher levels of demographic quota control did not increase data accuracy. Moreover, although some model-based sample selection approaches were found to reduce bias (improve proximity to benchmarks), there was no one method that provided universal improvement of the representativeness across all benchmarks that were used to access accuracy. Instead, the best method appears to be the patchwork quilt of methods which indicates that some methods prevail on some of the benchmarks and others prevail on the others. This may be because some of the models are balanced against the benchmark themselves or their correlates.

Capitalizing on Passive Data in Online Surveys

Tobias B. Konitzer, *Stanford University*
David Rothschild, *Microsoft Research*

Over the last few years there has been a dramatic decrease in the cost of opt-in, non-representative surveys. This has led to a massive increase in the quantity of available public opinion data, but it is less clear how accurately the data describes public opinion of the target population. One side of the debate has provided ample evidence that given the correct set of statistical adjustments, this data can yield meaningful estimates of population sentiment. In particular, research has shown that Bayesian post-stratification can yield robust estimates if the data fulfills some baseline criteria. However, this requires collecting a host of political and demographic variables that might be expensive to come by, even in more cost-effective opt-in surveys. Especially because demographic prompts can worsen non-response, both in that they lengthen surveys and in that they require revealing more sensitive information that might be seen as intrusive. This proposal explores how accurately these background variables can be collected passively in online surveys, without asking for them directly. First, demographics can be imputed reasonably well from Census data if Census block group of the

respondents can be determined by passive information collected with the actual survey responses. Second, modern polling platforms have means to match other information of the respondents, such as number and nature of installed apps if the survey is taken on a smart phone, which can help predict political-predispositional variables such as political interest and knowledge or news exposure. In this research, we conduct a survey on Pollfish, a polling company that maintains a panel of users recruited via Smart Phones, and assess how accurately we can predict demographic and political variables from passively collected data alone. We show that researchers can safely avoid actively collecting these variables crucial for statistical correction of non-response biases in non-probability surveys.

Sample Adjustments in Nonprobability Online Samples

Alan Roshwalb, *Ipsos Public Affairs*
Zachary Lewis, *Ipsos Public Affairs*
Robert Petrin, *Ipsos Public Affairs*

The qualities of online opt-in panels have been well-discussed in the context of properties of non-probability samples, limited rate of coverage of target populations, cooperation rates, and even in the context of mode effects to name a few. Adjustment strategies have looked at standard raking weights, propensity weights and propensity score matching. Auxiliary variables for adjustment have included standard demographics, technology adoption, and other measures. Most presented approaches do not seem to include an underlying strategic model that explains why the adjustment should work and to direct the adjustment approach. In this paper, we propose an attitudinal-behavioral consistency model to suggest and direct adjustment strategies. We will apply this model to attitudes and behavior towards influenza vaccinations. Other examples may be explored as well.

Refining ABS

Consistency and Accuracy of Undeliverable Codes Provided by the U.S. Postal Service: Implications for Frame Construction, Data Collection Operational Decisions and Response Rate Calculations

Kristine Fahrney Wiant, *RTI International*
Joe McMichael, *RTI International*
Joe Murphy, *RTI International*
Katie Morton, *RTI International*
Megan Waggy, *RTI International*

When survey mailings are returned as "undeliverable" in the US, we rely on information codes provided by the US Postal Service (USPS) to determine the eligibility of a sampled household. Although little is published about the consistency and accuracy of USPS undeliverable codes, they are often accepted at face value and used to make operational decisions about further contact and the ultimate disposition of a sample case. For the Residential Energy Consumption Survey (RECS), we examine the quality of the USPS's application of 8 different undeliverable codes. First, we review USPS codes from 3 mailings for the self-administered RECS National Pilot sent within a 12-day period to all 9,650 sampled households. We are particularly interested in codes that influence whether a case is disposed as eligible or ineligible. Next, we compare USPS codes to field-verified status collected from the 2015 RECS main, CAPI field study. Then, we assess a model to predict the undeliverable status of a case using auxiliary data that enhances known characteristics from the sampling frame. Finally, we discuss the implications

for operational tracking and decision-making on sample cases, the use of specific undeliverable codes in determining case eligibility, and the construction of sampling frames.

Evaluation of the National Survey of Children's Health Survey Design and Future Directions

Ronaldo Iachan, *ICF International*
Jason Fields, *U.S. Census Bureau*
Reem Ghandour, *Health Resources & Services Administration*
Echo Wang, *DS Federal*

This paper will draw lessons from the National Survey of Children's Health (NSCH) pretest, conducted by the Census Bureau, and from the evaluation conducted by the ICF team. This paper will describe a review conducted by ICF of the survey methodology planned for the NSCH. The redesigned multimode survey has the additional aim of combining two surveys of children conducted by the Health Resources and Services Administration's Maternal and Child Health Bureau. The integrated survey, the National Survey of Children's Health (NSCH), has several unique challenges as it screens for households with children and oversamples children with special health care needs. Guided by the Total Survey Error framework, we will assess the methods planned for the NSCH in their effectiveness in reducing coverage biases and the potential for non-response bias. We will also review the design's effectiveness of the two-phase stratified sampling design in maximizing precision, i.e., in minimizing survey variances for key subgroup estimates. In addition, we will assess approaches which may incorporate information about presence of children and internet access, linked from external data at the tract level and at the household level. These approaches can help the sampling design target likely eligible households and also assist the adaptive design reach out to web-enabled households in different ways. We will conclude with recommendations for the national production survey as well as ideas for similar surveys transitioning to dual mode, address-based sample surveys.

Should We Always Use the Telephone Numbers Matched to an ABS Sample?

Lawnzetta T. Yancey, *Nielsen*
Vrinda Nair, *Nielsen*

Nielsen has been using address based sampling (ABS) since 2008 for our TV diary survey. ABS replaced RDD because it offered better coverage of the population and it increased our ability to use a multi-mode survey design (both mail and phone). The residential telephone numbers appended to our ABS samples are from large consumer databases and the age of that data can vary depending on the source. In recent years, the ability of our vendor to match telephone numbers to the ABS sample is decreasing with the increase in cell phone only households and privacy concerns. Addresses with a telephone number are recruited by phone to participate in our survey. Unfortunately, contact rates are declining, in part to unusable telephone numbers. We normally find that ~40% of the 3.6 million phone numbers Nielsen phones in a year are categorized as unusable (not in service, out of order, or no answers) during phone recruitment. New phone screening technologies (CellWINS) from our vendor, Marketing Systems Group, Inc., has given us an opportunity to pre-screen the telephone numbers matched to our sample, thus increasing the efficiency of phoning. In the February 2016 TV diary survey, to further improve our sample performance, we are investigating removing the unusable phone number linked to the address and mailing the household a pre-pack instead. The pre-pack is provided to our ABS sample where we do not have a phone number to collect household demographics and a telephone number. Our goal with this test is to identify the shift in household characteristics captured in the

respondent data. Does it add more desired demographics to our sample? Will these homes participate at rates similar to the addresses never matched with a phone number? Will response rates increase enough to allow us to reduce our overall sample sizes?

Conducting a Telephone Survey Using an ABS Sample: A Case Study of the California Health Interview Survey

Jennifer Kali, *Westat*
Ismael Flores Cervantes, *Westat*

Geographically targeting samples in small areas is difficult in telephone surveys that include a cell phone component. There have been several solutions practitioners have used to increase the efficiency of targeting telephone surveys to small areas. One solution is the use of Address Based Sampling in which the sample unit is an address allowing for very precise geographic targeting. Because the sample unit is an address, an ABS frame is best suited for mail surveys. In some surveys, the flow of questions and skip patters in the instruments are too complex to be conducted on paper. Thus, there has been interest in using ABS as a way to recruit respondents for telephone surveys. To do so, a telephone number must be obtained for the addresses sampled via ABS. A telephone number can either be matched to the sampled address or obtained via a letter is mailed to the sampled address requesting respondents to provide their telephone number. The California Health Interview Survey is a dual-frame RDD telephone survey of the population of California. In 2013-2014, ABS was used to supplement the RDD sample in one rural county in which geographic targeting for the cell phone sample required a large screening effort. This presentation describes the CHIS experiences with using ABS to conduct a telephone survey and examines the differences between the ABS and RDD samples. Methods to obtain telephone numbers, data collection outcomes, including response rates, disposition codes, and accuracy of the telephone matching, and estimates of health and demographics are compared between the RDD and ABS samples.

Using Auxiliary Data to Increase Efficiency of Sampling Rental Units in Metropolitan Regions

Randal ZuWallack, *ICF International*
Joshua Brown, *ICF International*
Thomas Brassell, *ICF International*
Richard Williams, *Vermont State Housing Authority*
Ron Dion, *Housing Authority of the City of Alameda*
Vanessa Cooper, *Housing Authority of the City of Alameda*

This presentation builds on our analysis of using auxiliary data to increase the efficiency of locating renters to conduct Fair Market Rent (FMR) Surveys for local public housing authorities (PHAs). In 2014, we analyzed the use of a rent flag and geographic rental density to oversample addresses for two rural counties in Vermont. This study expands the analysis to a large metropolitan area, the Oakland-Hayward-Berkeley, CA metropolitan division and a small metropolitan area, the Burlington-South Burlington MSA. In 2014, we found that 95% of the addresses flagged as "owned" were indeed owners. While the percentage of renters among the flagged "rented" units was smaller (64%), we gained efficiencies by under-sampling the owned cases. We estimated cost savings of 20%-30% based on the optimal sampling allocation. The Oakland metropolitan area has a higher renter population (41%) than the rural Vermont counties (30%), but similar turnover in rental housing--26% of Oakland renters did not live in the same residence one year ago compared to 28% in the rural Vermont counties. The

Burlington metropolitan area has similar rental rate to the rural areas (32%) but higher turnover in rental housing (35%). Given the different nature of the housing environment in these areas, we repeat our analysis to determine whether the auxiliary information provides similar efficiency benefits for sampling rental units. We hypothesize that a greater number of records will be flagged in the metropolitan areas, but the accuracy of the flag could be compromised with higher housing turnover.

Measuring Attitudes on Abortion – Time for Another Look

Panel organizer: Kate Stewart, *ConwayStrategic*

The traditional abortion question which has been asked for decades would have us believe that the country has barely moved on this issue since abortion was legalized in 1973. When attitudes on abortion are reported in the media, we most often hear that views on abortion are "closely divided" and have been "stable" for decades. Given the level of debate around abortion policy nationally, the dramatic rise in state laws on abortion access, threats to shutdown the federal government over Planned Parenthood funding, and an impending Supreme Court case, it is time we take another look at how we measure abortion attitudes. Most major pollsters ask Americans to indicate whether their beliefs align with one of the following four categories: Abortion should be legal in all cases; abortion should be legal in most cases; abortion should be illegal in most cases; or abortion should be illegal in all cases. Some surveys also ask whether individuals identify with the labels "pro-choice" or "pro-life" and increasingly have found that sizeable segments, in particular younger Americans, identify as both. Public Religion Research Institute recently released a survey and found nearly half of millennials (age 18-35) identify with both "pro-life" and "pro-choice" labels (27 percent) or identify with neither (22 percent). Only about one-quarter identify exclusively as "pro-choice" (27 percent) or as "pro-life" (25 percent). This panel will examine the conventional abortion measurement tools, raise questions about whether we need to review how we measure attitudes on abortion, and begin to provide ideas for ways to create tools which more accurately tap into the complex and, at many times, conflicting attitudes the public holds on abortion. This panel will bring together researchers who have been studying attitudes on abortion along with journalists who cover and report on these issues.

Context Matters: Assessing the Validity of Common Survey Questions on Abortion
Jill Mizell, *Center for Reproductive Rights*

Jill Mizell, AAPOR member since 2008, holds an M.S. in Applied Social Research from Hunter College and an M.S. in Statistics from Baruch College. The questions used to measure public opinion on a woman's right to safe and legal abortion care are reliable, but are they valid? Literature has shown that attitudes on a woman's legal right to abortion vary widely depending on context. Usually these claims are based on questions that include language around timing and gestation of a woman's pregnancy, her primary reason for obtaining abortion care, or both. Two studies shed additional light on other potential factors at play – differentiating personal beliefs from what should be generally true for society, and establishing the status quo. The survey experiments used a split sample design to test the validity of common questions measuring abortion attitudes. One experiment found that adding language that primes people to think about social policy, rather than what they would personally choose in their own lives, changed how people answered this question. Another experiment showed that establishing the current legal status and parameters – that the Supreme Court has ruled abortion before viability is legal – also affected how people answered the question. Together, this research builds on the

literature that suggests that asking abortion policy questions in the abstract yields statistical artifacts, and that research should seek to improve the validity of the measures we use to enhance our understanding of the complexity of attitudes about a woman's legal right to safe abortion care.

It's Time to Retire the Legality Question of Abortion
Tresa Undem, *Perry/Undem*

Tresa Undem is a partner at PerryUndem Research/Communication. Her clients include the Robert Wood Johnson Foundation, The Henry J. Kaiser Family Foundation, and the National Institute for Reproductive Health. Ms. Undem will present quantitative research that raises questions about the accuracy and relevance of standard polling questions on abortion. These questions tend to ask about legality of abortion in "all cases, most cases, just in rare cases, or never legal." Results for decades show a divided public. Over the past year, Tresa's research has found that these typical measures on the legality of abortion, used by her firm and many others, are at best insufficient, and at worst, may be inaccurate and misleading. For example, among those who believe abortion should be illegal except in rare cases, 53% do not want to overturn a woman's constitutional right to abortion; 45% believe the recent restrictions on abortion are going in the wrong direction; and 71% would support a friend or family member who decided to have an abortion.

Conflicted Hearts and Minds: In-depth Qualitative Insights on Abortion Attitudes
Amy Simmon, *Goodwin Simon Strategic Research*

Amy Simon is a Founding Partner at Goodwin Simon Strategic Research, and has worked on candidate, ballot measure and public education campaigns at the federal, state and local level in over 40 U.S. states, as well as internationally. Ms. Simon has a special expertise in conducting innovative qualitative and quantitative research on socially controversial and emotionally complex issues. While numerous quantitative measures on abortion attitudes currently exist—including the ubiquitous "four-way" abortion question—our extensive qualitative research reveals a great deal of complexity behind those quantitative measures. For most people, the subject of abortion triggers a wide range of emotions, judgments, and other considerations that standard abortion measures are unable to capture in a deep and meaningful way. In particular, we find multiple dimensions at play as people grapple with issues of legality, morality, family, responsibility, and the "context" of abortion access—either for one's self/partner, for other individuals, or for women (and families) in general. This presentation will explore the various reasoning structures we have identified in our qualitative research utilizing individual in-depth interviews and focus groups. Notably, this research has been conducted across a diverse array of audiences, including white women, white men, African-American women, African-American men, Latina women, and Latino men.

Creative Ways to Think About Abortion and Public Opinion
Peyton Craighill, *U.S. Department of State*

Peyton Craighill is the former polling manager of the Washington Post and future director of polling for Europe/Eurasia at the U.S Department of State. A stable trend can discourage pollsters to revisit a question for which they feel they know the answer. Opinions on abortion, specifically whether it should be legal or illegal, have been durably consistent in Washington Post-ABC News polls and other public polls. Is there something going on that we are missing? I will explore alternatives to traditional analysis for this trend question, as well as explore other public sources on abortion attitudes. Finally, I'll look at how you change a question to make it more au courant without sacrificing valuable trend.

Covering the Abortion Issue and the Need for Better Measurement of Public Attitudes
Sarah Kliff, *Vox*

Sarah Kliff is a senior editor at Vox, where she oversees policy coverage and writes about health care. She has covered reproductive health for eight years both at Vox and other outlets, including Newsweek, Politico, and the Washington Post. Sarah will discuss the discrepancies she's seen in her work as a reporter between how pollsters ask questions about abortion — and what the public actually thinks about the issue. In interviewing numerous Americans on the topic, she's found that their issues aren't nearly as clear as the "pro-life" and "pro-choice" labels. Instead, most hold views that are much more nuanced than that, living in a grayer area in between the two poles. Sarah's presentation will focus on how better polling data could improve news coverage of the abortion issue, generating more nuanced stories that reflect more accurately how Americans think about this important and persistent policy issue.

May 14, 2016
Concurrent Session H

Optimizing User Experience for Survey Instruments on Smartphones

Panel organizer: Jennie Lai, *Google*

The phenomenal growth of smartphone users has reshaped the research landscape for survey researchers in recent years. Almost two-thirds of American adults (64%) now own a smartphone compared to just over a third (35%) in 2011 when Pew Research Center first conducted its survey on smartphone ownership in the U.S. Survey researchers have since leveraged the smartphone device as survey instruments for ad-hoc or continuous data collection through its browser or apps for self-report or passive measurement. The possibilities are endless but not without challenges. How does this fast-evolving landscape of smartphone ownership and varying levels of experience with its features, coupled with survey respondent device choice, affect the data collection process? Mick Couper's Usability Evaluation of Computer-Assisted Survey Instruments (2000) first reported on usability testing for interviewer-administered and self-administered surveys by respondents. Its foray into user-centered design for survey instruments on computers has inspired this panel on optimizing user experience for survey instruments on smartphones more than a decade later. The movement of user-centered design, i.e., design driven by user needs and preferences, for survey research on smartphones is imminent. This panel offers examples of usability evaluation using eye-tracking method for surveys on smartphones as well as cognitive walkthrough of a social media app as a survey instrument. In addition to these evaluative techniques for survey design, it also introduces case studies of leveraging in-context survey on user experience for mobile products at scale and passive measurement of behavioral data and other data formats beyond self-administered surveys. The discussion of this panel will be led by Mick Couper on the changing research landscape of survey instruments from computer to smartphone and its implication on methods for usability evaluation and data collection.

Eye Tracking for Surveys on Smartphones

Jennifer Romano-Bergstrom, *Facebook*

In survey research, we often focus on cognitive testing to ensure respondents properly understand instruction, questions, and response options. Usability testing, which focuses on how respondents use the surveys, is often forgotten. Additionally, when usability testing is conducted, mobile is often forgotten. Yet, mobile is often the first and only device many consumers have to connect to the Internet, especially in emerging countries and in areas with low socio-economic populations. To ensure a product is made for all users, it is important for companies to consider designing and creating for mobile as well as desktop. Understanding what users look at while using a product, is not as difficult as it once was. Modern technology has made this eye-tracking technique quite simple and affordable.In this talk, you will learn about modern eye tracking and how it can help you to understand your users. We will discuss recent usability testing projects that including eye tracking and revealed important findings that led to changing the appearance of Facebook products and ultimately improved the user experience. Attendees will learn about the importance of conducting usability testing during survey development, pros and cons of using eye tracking with mobile devices, and what the different types of data and visualizations can tell you.

Designing New Data Capture Method: Usability Study of Instagram App as a Data Collection Tool

Jennifer Kelley, *University of Michigan*
H. Yanna Yan, *University of Michigan*

With the explosion of social media usage, survey organizations are looking for innovative ways to leverage social media to either supplement, or in some cases replace traditional survey methods. Using data from social media platforms is appealing for many reasons, including the relative ease of extracting 'big' data, capitalizing on users existing knowledge of how to use the tool, and the real-time, 'natural' behavior aspect of the data. While there is a growing segment of survey researchers exploring and using Facebook and Twitter for data collection, there has been little exploration and use of Instagram for data collection. Instagram's primary feature is sharing photos. Photos, especially in diary surveys or product surveys, can add a rich layer of data to a study. While Instagram is a growing social networking platform with Pew's 2015 Social Media Update reporting that usage has doubled from 13% in 2012 to 28% in 2015, nearly two-thirds of the population are non-Instagram user. The question becomes, can non-Instagram users be trained to use Instagram as a data collection tool? One logical first step is to conduct a usability test to understand how both Instagram users and non-users interact with Instagram. This presentation will discuss the usability study that was conducted to inform a pilot for Feeding America to learn how individuals in need of food assistance find Feeding America's food agencies and clients' food preferences. The pilot's design was to use Instagram to collect various types of data: image, geographic location, general text and hashtags. The usability testing results contributed to the success of the Spring 2015 pilot study with five Feeding America's partner agencies. We will discuss the insight we gained on potential technical and methodological issues, including the training of data collectors and the data capture design, using Instagram as a data collection tool.

User Experience Considerations for Contextual Product Surveys on Smartphones

Aaron Sedley, *Google*
Hendrik Müller, *Google*

As use of smartphone-based tools & services broadens and deepens, such products should be continuously evaluated and optimized to meet users' needs. In-product surveys are one way to gather attitudinal and user experience data at scale, in context of actual experiences. However, given space constraints, OS variance, product design differences, and app vs mobile web options, launching contextual surveys on smartphones requires considerable attention to several user experience aspects that can also impact data quality. In this talk, we will discuss a variety of practical UX considerations for in-context surveys on smartphones, drawing on real-world implementation and experimentation for several of Google's most-used mobile products. In particular, we'll discuss issues such as when to trigger a survey, sampling across platforms (mobile vs desktop), invitations vs inline questions, survey length, question types, device size, screen orientation, survey interaction with the host product/app. We will also explore the effect of design and text variants on smartphone survey response rates and response distributions.

Beyond the Survey: Improving Data Insights and User Experience with Mobile Devices

Patricia Graham, *GfK Custom Research, LLC*
Gavin Lew, *GfK Custom Research, LLC*

Social science and commercial researchers rely on surveys to ask people questions so that attitudes and behaviors are inferred from self-reported responses. Yet response rates decline, frames are difficult to sample, people cannot remember depth of their digital behavior and attention spans diminish while survey research consumer need holistic information. Devices people use provide opportunities to address these persistent issues with less intrusiveness on experience. Roughly 30% of US web surveys are completed on a mobile device and these participants can tell us more than asking questions can provide! The authors present additional uses of mobile technology to provide information beyond what is traditionally gathered in surveys. We provide two case studies to illuminate not only what to do to gather these new forms of information, but also the mechanics of collecting the data. These case studies use multi-media and passive monitoring that will be shown to provide additional information about populations of interest in less intrusive and more engaging ways. The passive monitoring case study empirically illustrates the depth of behavior obtained to fully understand digital consumption by party ID married with attitudinal profiling. The industry will see how passively metered behavior can improve ability to understand people's thoughts and behaviors. Case two enumerates how tipping points with a visual record of web path and decision points with visceral in-the-moment reactions recorded using mobile multimedia data capture can advance insights. Providing our industry an understanding of data capture considerations we review practices that are used for passive metering of persons and 'in the wild' active user experience. We examine findings concerning sample representativeness and data quality, with an eye toward reducing the biases associated with mobile samples. Moreover, we assess the relevance definitions of sample compliance have on data quality.

Methods to Improve Data Quality

An Exploration of the Relationship between Usability Testing and Data Verification

Casey Langer Tesfaye, *Nielsen*
Vera Kurmlavage, *Nielsen*

Visual design is an important aspect of web survey development because people use the visual cues around a question to help them understanding the meaning and intent of the question. Usability testing helps us to understand how our designs are functioning and how our survey questions are interpreted. Usability testing also helps us to identify barriers to survey completion. Over the course of 14 months, Nielsen conducted iterative usability tests based on a web form and then developed and implemented design changes to address these fixes. For this study, we compare the results of a data validation study that was conducted on an early version of the form with the results of a data validation study (using the same questions) that was conducted after a significant number of usability improvements were implemented. The combination of quantitative and qualitative pretesting done on this set of questions affords us the unique opportunity to take a deeper look into the interaction between these different pretesting processes. By comparing data quality across these versions of the survey, mapping these differences back to the usability testing, and drawing on qualitative commentary that was collected in a series of in-depth interviews conducted after the validation studies, we will be able to take a closer look into the interaction of data quality and usability testing. We will explore the question: what would have happened had we released the survey without the extensive pretesting and how much did we really gain from our efforts? Ultimately, our goal is to produce a set of data-driven, practical recommendations for researchers who hope to maximize their usability testing efforts.

Parents and Teens: Proxy versus Self-report to Measure Teen Disability

Meredith Massey, *National Center for Health Statistics*

Who should answer questions about teen disability: teens or their parents? Parents are generally viewed as suitable proxies for their children since children may have difficulty answering questions due to age, illness, disability or literacy and because it is the parents who most often seek medical care and support services for their children. Therefore, many surveys that measure disability prevalence are designed to be administered to parents. However, older children, teens in particular, may be better able to evaluate their own difficulties especially in domains that are focused on internal processes such as memory or emotions. Using data from a cognitive interview study testing the UNICEF/Washington Group on Disability Statistics Module on Child Functioning and Disability, differences in the ways parents and teens answer and interpret questions on teen disability will be explored. The results of this study, which collected data from 40 parent-teen dyads (n=80), demonstrate that overall, parents and teens answered similarly. However, it was seen that the dyads were more closely aligned in their evaluations of questions categorized as 'behavior' questions than they were for those categorized as 'ability' questions. Some discordance in teens' and parents' responses was a result of their different interpretations of the questions. Further, for some questions, parents and teens had differing patterns of interpretation even when they answered similarly. Other factors that influence dyad agreement (such as age and gender), will also be explored. Finally, we will outline the implications of these findings on the design and administration of surveys measuring the prevalence of teen disability.

The Effects of Sending a Pre-notice Postcard in a CAPI-conducted Survey on Response and Cost

Allison Zotti, *U.S. Census Bureau*

During the past twenty years, the cost of computer-assisted personal interviewing (CAPI) government surveys has been increasing, while - simultaneously - many of their response rates have been decreasing (Malone, 2006). The need for techniques to increase response rates or decrease costs are necessary. This study examines one such technique, sending a pre-notice postcard to a CAPI-conducted survey. The goal of this study was to increase the response rate and decrease the cost by reducing the number of contact attempts needed to complete an interview on a sample address or case. The postcards include the name and contact number of the interviewer assigned to that particular case, which gives the respondent an opportunity to call their interviewer and respond or set up an appointment. This study analyzed data from 500 interviewers and 25,000 cases in the Current Population Survey (CPS), a government survey that is the nation's leading source for the unemployment rate and has recently experienced a drop in response rates (Census Bureau, 2006). Housing units are in a rotating 4-8-4 panel design, where they are in the sample for four months, fall out of the sample the next eight months, and are then back in sample for another four months. Field interviewers conduct the majority of the interviews, either in person or by phone. This study was conducted in three of the six field management offices (Chicago, Denver, and Philadelphia) in July, August, and October 2015. This design allowed for the determination of any potential geographic differences or additive effects from repeatedly receiving a pre-notice postcard. Mixed effect logistic models were fit to determine what characteristics, including receiving a postcard, predict a cases' propensity to respond. Count regression was used to determine what effect the postcard had on the number of contact attempts needed to complete an interview.

Benefits and Challenges of Using Language Preference Indicators to Tailor Mail Surveys by Language

HarmoniJoie Noel, *American Institutes for Research*
Chris Evensen, *American Institutes for Research*
Alison R. Huang, *American Institutes for Research*
Carol Wan, *American Institutes for Research*

Tailoring mail surveys in the language preferred by respondents is a best practice but language preference indicators are not always available on sampling frames. For mail surveys, alternatives include providing contact information in other languages on cover letters to enable people to request non-English surveys, which results in low up-take. Double stuffing two languages, usually English and Spanish, is expensive and inefficient. In the Qualified Health Plan Enrollee Experience Survey field test, a survey of people obtaining health insurance through the Affordable Care Act's Marketplaces, we obtained language preference on the sampling frame from their Marketplace application to send Spanish and Chinese questionnaires directly, which makes these data unique. As a result of stratifying our sample by language and tailoring the mail surveys by language, we had a substantial number of Spanish and Chinese respondents. This study explores the effects of using information about Spanish and Chinese language to understand reduction in nonresponse biases, the psychometric properties of survey measures, and reliability. We compared results with and without Spanish and Chinese speakers to see the effects of their inclusion in terms of response rates, internal consistency reliability of composite measures, and inter-unit reliability, which shows how well the measures distinguish between health plans. We assess overall data quality in terms of failed skips and item missing data. Response

rates were higher for Spanish and Chinese than English respondents. The benefits of including Spanish and Chinese respondents are: more variability in outcomes and more reliable estimates. However, there were more failed skips in the Spanish and Chinese data so they required more cleaning. Including non-English speakers had benefits in terms of the results and conclusions. Government agencies and other sources of sampling frames (i.e. health insurance companies) should include and use language preference to send materials in respondents' preferred language.

Goodbye Humans:
Robots, Drones and Wearables as Data Collectors

Using Drones for Household Enumeration and Estimation

Safaa R. Amer, *RTI International*
Mark Bruhn, *RTI International*
Karol Krotki, *RTI International*

In many countries, the census of population and dwellings is the basis for implementing household surveys. In between censuses, estimates are updated using projection models. Censuses are large- scale, resource-intensive efforts that in many countries suffer from under-coverage and other sources of error. In addition, projection models are based on specific assumptions which might not apply. In an attempt to reduce error in projection models and provide planning tools which can reduce the cost of dwelling and household census, we conducted a case study on the use of drones to improve estimates of dwellings and households in between census and provide planning tools for upcoming household survey not to mention the next census data collection. In this paper, we discuss the use of drone imagery combined with point cloud technology to generate 3D images of buildings and help in the enumeration of household without having to use field staff to do a traditional listing. We also describe a system for extracting information from these images to provide more accurate information for updating dwelling and household counts. We investigate the use of the enhanced imagery to support the planning of household surveys and census operations. We compare the drone imagery with information available from aerial satellite photography and in some limited cases the result of traditional listing operations. For various logistical and legal reasons, the drone experiment was limited to the Galapagos Islands but we expect that with increased acceptance of this new technology, survey researchers will be able to use drone technology for the purpose of household listing in many areas throughout the world.

Robots as Survey Administrators: Adapting Survey Administration Based on Paradata

Ning Gong, *Temple University*
Nina DePena Hoe, *Temple University*
Carole Tucker, *Temple University*
Li Bai, *Temple University*
Heidi E. Grunwald, *Temple University*

Background: The use of robots has emerged within healthcare and education settings to provide training and opportunities for social interaction. Robotic systems have been developed for children with Autism Spectrum Disorders (ASD) for education and training with early success. Patient reported outcomes (PROs) have become common and provide critical information about a person's health, feelings, internal experiences and symptoms. In this project we develop a robot system capable of collecting paradata that is used to adapt the timing and content of the administration in real-time for

child health surveys. Methods: The NAO Robot (Aldebaran, France) was programmed to adaptively administer health surveys to children. The robot stated the question, recorded the response and survey paradata, as well as ambient light, noise, the respondent's facial expression, voice volume, eye contact, and physical proximity. The robotic system is designed to integrate with physiologic data (heart rate, limb movement) collected in real-time from a Microsoft band. A communication protocol is created to transmit the data to a cloud-based server and to the robot for post and real-time data processing. Pilot data was collected during system development on young children and adolescents with ASD. Results: The survey administration is adapted based on the child's request for breaks and a rule-based strategy for adapting the content and timing of administration. It is developed based on paradata of the child's physiologic responses and response patterns. The parents, clinicians, and child respondents reported that the robot system survey was enjoyable and provided more ecologically valid interactive administration. Conclusion: This project is unique in its integration of robotic survey adaptive administration based on a respondent's behaviors and physiologic responses. The paradata approach through robot technology opens vast opportunities to advance PRO research and increase the validity and reliability of PROs from respondents with socio-emotional/behavioral or cognitive challenges.

Wearables: Passive Media Measurement Tool of the Future

Adam Gluck, *Nielsen*
Leah Christian, *Nielsen*
Jenna Levy, *Nielsen*
Victoria J. Hoverman, *Nielsen*
Arianne Buckley, *Nielsen*
Ekua Kendall, *Nielsen*
Erin Wittkowski, *Nielsen*

Wearable technology includes a myriad of devices that possess advanced functionality, connect to the Internet (typically via smartphone), and are usually designed to be worn on the body persistently (Endeavour Partners, 2014; Khan & Marzec 2014). As researchers become increasingly focused on more passive ways to measure people's behavior, wearable technology provides a comprehensive method of data collection. The appeal of wearable technology for research purposes lies in the fact that it gives up-to-the-minute data without distracting the user from other responsibilities and activities. Users expect wearable technology to be multi-functioning, with a fast response time, and very little interaction between the user and device (Cognizant, 2014). Nielsen uses a variety of devices to passively measure media consumption, including wearable meters that measure media exposure to encoded television and radio signals. People are recruited and asked to wear or carry the meter wherever they go throughout their day. The current form of the wearable meter, a pager-like device, is a significant barrier to compliance particularly for key subgroups such as young women and children. Our research explores challenges with carrying or wearing a meter, people's interest and use of wearable devices, preferences in form and additional functionality, and willingness to wear for sustained periods of time. The goal is to evaluate whether people can better integrate a more wearable meter that leverages these new designs into their everyday lives, improving overall measurement and compliance. We will present findings from an iterative, mixed method approach to exploring alternative form factors, including an extensive literature review, a survey and focus groups with former panelists, a nationally representative survey of the U.S. population, and focus groups with field staff who interact with panelists.

QR Codes for Survey Access: Is It Worth It?

Laura Allen, *The Gallup Organization*
Jenny Marlar, *The Gallup Organization*

With the emergence of new technologies, survey methodology must adapt appropriately to help motivate responses and ensure progression of the field. Quick Response (QR) codes provide a potentially useful tool in mobile access, offering instant access to information that is nearly limitless in possibility. This study examines the use of QR codes as a point of mobile access to web-based surveys. Research has shown that QR codes can provide a simple and effective way to distribute information that is customizable to fit the needs of the project or task at hand (Lombardo, Morrow, & LeBur, 2012). Though it may seem a quick and easy method that is in vogue with the mobile movement in survey methodology, there are also some limitations to consider. It is important that new methodologies do not deter some demographics for the sake of appealing to a small amount of respondents. Studies have shown that few respondents knew what a QR code was, let alone had the required application to read the code (Lane et al., 2012).

Gallup is fielding a general population survey of economic issues with an embedded a multi-mode experiment. Using an ABS sample, respondents were randomly assigned to different treatments: (1) mail survey with option to also complete online by typing a URL into a browser, (2) mail survey with options to complete online by typing a URL or scanning a QR code, and (3) mail survey with option to complete online by scanning a QR code. These experimental groups will be compared on a variety of factors including overall response rates, number of responses by mode option within treatment group, demographics, and substantive results. Preliminary response rates show no significant difference between the treatment groups with and without QR codes.

Comparing Youth's Emotional Reactions to Traditional vs. Non-traditional Truth Advertising Using Biometric Measurements and Facial Coding

Jessica M. Rath, *Truth Initiative*
Morgane A. Bennett, *Truth Initiative*
Mary Dominguez, *Truth Initiative*
Elizabeth C. Hair, *Truth Initiative*
Donna Vallone, *Truth Initiative*
Naomi Nuta, *Nielsen Consumer Neuroscience*
Michelle Lee, *Nielsen Consumer Neuroscience*
Patti Wakeling, *Nielsen Consumer Neuroscience*
Mark Loughney, *Turner Broadcasting*
Dana Shaddows, *Turner Broadcasting*

Truth®, a national anti-tobacco marketing campaign, was shown to be successful in changing youth's attitudes towards tobacco. Integration of truth® messaging with TV content popular among youth is another method of reaching the target market. Biometric and facial coding measurements assessed differences in emotional responses to a truth® ad ("Left Swipe Dat") and an Adult Swim TV network custom integration ad. Participants included 375 Adult Swim viewers, ages 16-21, from 5 U.S. markets. Participants were randomized into: 1) "Left Swipe Dat" ad only, 2) integration only, or 3) both ad and integration. Biometric measures of heart rate and skin conductance were passively collected with Nielsen Consumer Neuroscience's (NCNS) technology while participants viewed the ad content. NCNS' patented algorithms identified moment-by-moment levels of emotional engagement, while facial coding captured expressed emotion through discrete facial expressions of positive, surprise, or

negative/confusion. Participants viewing "Left Swipe Dat" in isolation exhibited high levels of engagement for 29% more time than those who viewed the integration alone (45% vs 16%, respectively). A priming effect was evident when participants viewed "Left Swipe Dat" before the integration, resulting in an increase in amount of time spent highly engaged to 72% when viewing the integration. Additionally, prefacing the integration with "Left Swipe Dat" led to positive expressed emotion at the mention of "left swipe" in the integration. This positive expressed emotion to the anti-smoking concept of "left swipe" is not present when the integration is viewed alone. These methods allow researchers the unique opportunity to better understand campaign success by analyzing unconscious reactions to ad content. Biometrics and facial coding measures provides a deeper understanding of audiences' emotional responses to ad content, which can complement data collected through self-report. Understanding the effect of priming audiences with related content can aid in the development of more effective campaign materials.

Methods Research in Online Surveys

Influences on Item Response Times in a Multinational Web Survey

Benjamin Phillips, *Abt SRBI*
Stanislav Kolenikov, *Abt SRBI*
Elaine Howard Ecklund, *Rice University*
Allison Ackermann, *Abt SRBI*
Alec Brulia, *Abt SRBI*

We model time to respond in web surveys of members of biology and physics departments in French, Italian, Turkish, and U.S. universities and research institutes to understand factors associated with time to respond to survey items in a cross-national, multilingual context. Our findings identify points at which respondent attention diminishes, providing guidance on optimal length of item stems, response options, and survey length for similar populations.The Rice University Religion among Scientists in International Context (RASIC) survey included measures of time to respond. The survey provides a rich source of material. Respondent-level measures include biographical data including age, academic rank, language of choice (the survey was offered in the native language and English in non-U.S. locales), and country of origin. Item-level measures include length of item, reading difficulty, topic, number of responses, and position in survey. Paradata include accumulated time spent on the survey, and time of day. We find inflection points beyond which we see satisficing in the form of diminished respondent attention for the following factors: number of words in item stems, time from start of the survey. Differences in inflection points by language of survey are analyzed by respondent country of birth to understand variations for nonnative speakers. Variations in time of response for sequence of item in instrument (controlling for time), question type, time of day, day of week, and academic rank are also seen. No effect is found for reading grade level, number of response options (controlling for words in response options), gender, inclusion of a "don't know" option, scientific discipline, or restarting the survey. A hierarchical cross-classified model is used for analysis. Implications of these findings for questionnaire design are discussed.RASIC data collection was funded by the Templeton World Charity Foundation, grant TWCF0033.AB14, Elaine Howard Ecklund, PI, Kirstin RW Matthews and Steven W. Lewis, co-PIs.

Longitudinal Online Ego-centric Social Network Data Collection with EgoWeb 2.0

Alerk Amin, *RAND Corporation*
David Kennedy, *RAND Corporation*

Social network analysis is a key methodology in modern sociological research. Research has shown that social networks have a large effect on respondents' behaviors and attitudes, including health behaviors, risk taking, and many others. Longitudinal tracking of the changing social network of respondents provides new insights into how people change over time. However, collecting data about social networks, particularly in online surveys, can be burdensome for respondents. Developing technologies to ease respondent burden is crucial to expanding this research. EgoWeb 2.0 is an open-source platform for conducting online and face-to-face surveys to gather ego-centered network data. It allows researchers to collect data not only about the structure of the social network, but also the characteristics of the alters, the relationship between the ego and each alter, and the relationships between alters. It includes options to present network visualization immediately after collection of network structural data to elicit additional data based on a respondent's evaluation of the network display. For longitudinal studies, networks are pre-loaded to greatly reduce respondent burden and underreporting of network contacts. The EgoWeb 2.0 tool will be presented, along with longitudinal studies that have used EgoWeb for online collection. The results from the Context of Adolescent Risk Behavior in Networks (CARBIN), a longitudinal study of substance abuse from three public schools in Illinois, will be presented. The results of a study with a nationally representative probability-based sample from the American Life Panel will also be presented. The insights from these projects will demonstrate the feasibility and data quality from longitudinal online social network surveys.

Taking it Online: Creating a Radio Diary for the Online Age

Stephanie Stalinski, *Numeris*
Natasha Arzumanian, *Numeris*

Twice per year, Numeris conducts radio ratings "sweeps" in Canada. Potential households are contacted by phone and enumerated to determine basic demographic composition. For households agreeing to participate, every respondent aged 12 or older is mailed a paper diary in which they record their radio listening for a one week period. Numeris then compiles these data into the ratings which are used as currency in the Canadian marketplace. Drawing on previous experience in this area, Numeris has spent the past two years redeveloping and testing an online, fully mobile-compatible version of its radio diary. A small-scale field test resulted in good uptake of the online option (37%) but with a marked reduction in return rates for online compared to paper respondents. Following this test, a number of improvements were made to the online diary which made it easier to access and incorporated electronic reminders for online respondents. Subsequently, a large-scale field was conducted in different test markets. This test found high uptake of the online option (46%) with return rates for online respondents almost the same as for paper respondents. While online respondents tended to be younger than paper respondents, there were no consistent demographic differences between the test group (online + paper respondents) and the production sample (paper only). A second large-scale field test was recently conducted in the same test markets as the first large-scale test. The results of this test will provide insight into the consistency in uptake of the online tool, return rates, and listening data across survey periods.

Non-response bias for Estimates Sexual Assault and Sexual Misconduct on a Large Campus Climate Survey

David Cantor, *Westat*
Reanne L.M. Townsend, *Westat*

An important question related to self-report surveys of rape and sexual assault (RSA) is the extent non-response bias affects incidence and prevalence estimates. One argument is that non-victims will be less likely to respond to the survey because of lack of interest in the topic. An alternative hypothesis is that those most affected by RSA will not respond because of a reluctance to answer questions. This presentation will report on an assessment of non-response bias in a large campus climate survey administered across 27 Institutions of Higher Education (IHE) coordinated by the Association of American Universities (AAU). The aggregate response rate was 19%, but this ranged from 7% to 53% across the 27 IHE's. Analysis to date indicates that if bias exists, non-respondents tend to be non-victims and that the size of the bias is relatively small (e.g., several percentage points). These analyses have focused on internal comparisons between early vs late responders, as well as between incentive and non-incentive groups. One weakness of these analyses is they compare groups with relatively low overall response rates. For example, the difference in response rate between the incentive groups is 9 percentage points (15% vs. 24%). This paper will extend these analyses by comparing the AAU estimates to several other surveys that have used similar measures of campus climate but achieved significantly higher response rates. The comparisons to other surveys provide an external benchmark to assess the magnitude of the bias when the response rate is significantly higher. For example, one comparison will be with a survey that achieved a response rate close to 70% and was administered to the same student population as the AAU survey, but earlier in the school year. Comparisons will also be made to several other surveys with response rate ranging between 40% to 80%.

Impact of Field Period Length in the Estimates of Sexual Victimization in a Web-based Survey of College Females

Marcus E. Berzofsky, *RTI International*
Kimberly C. Peterson, *RTI International*
Bonnie E. Shook-Sa, *RTI International*
Christine Lindquist, *RTI International*
Christopher Krebs, *RTI International*

When administering a survey on a sensitive topic, such as sexual victimization, one needs to be concerned about the potential for bias due to the length of the survey's field period. If persons who have a greater interest in the survey topic (e.g., because they are victims) are more likely to respond quickly then a short field period may lead to upwardly biased results. However, a long field period may negatively impact estimates when there is a fixed starting point for the reference year (e.g., beginning of the academic year) because the reference period for early responders is shorter than the reference period for late responders. The Campus Climate Survey Validation Study (CCSVS) Pilot Test , sponsored by the Bureau of Justice Statistics and the Office of Violence Against Women, was a web-based survey administered at nine colleges interested in measuring the prevalence and incidence of sexual victimization among undergraduate students during the 2014-15 academic year. The survey, administered at the end of the Spring 2015 semester, was in the field for approximately 60 days at each school even though almost all schools achieved their targeted sample size within 28 days. In this paper

we present a comparison of the estimates for key sexual victimization outcomes based on three different field periods. We found that early and late responders did not differ for the key outcomes of interest. We discuss how the use of incentives and other strategies in the CCSVS may have contributed to mitigating potential bias in terms of how long it took victims and non-victims to decide to participate in the survey. Furthermore, we look at how response rates varied by the different field period lengths and the impact that shorter field periods would have had on design effects after adjusting for nonresponse.

Sexual Assault on College Campuses: Key Findings from a KFF/Washington Post Survey of Students

Elizabeth Hamel, *The Henry J. Kaiser Family Foundation*
Bianca DiJulio, *The Henry J. Kaiser Family Foundation*
Scott Clement, *The Washington Post*
Peyton Craighill, *U.S. Department of State*

Sexual assault on college campuses has gotten national attention recently, with Title IX investigations, a White House Task Force on the issue, and broad media coverage. To find out what students themselves think of the issue and their experience with it, The Henry J. Kaiser Family Foundation and The Washington Post conducted a nationally representative telephone survey in early 2015 of current and recent full-time, undergraduate college students living on or near campus ages 17-26 on sexual assault – one of the most comprehensive surveys to date on the topic. The survey finds that twenty percent of young women who attended college during the past four years say they were sexually assaulted. Nearly four in ten current and recent students say sexual assault is a problem on campus, ranking higher than cheating in academic, sexism, and property crime, but falling behind alcohol and drug use. The presentation will explore the experiences of survivors themselves, as well as students' views more broadly including the scope of the problem, prevention strategies, commonly held attitudes or myths about sexual assault, and school administrations' handling of incidents. The presentation will also highlight the process of conducting a survey on such a sensitive topic, including reviewing best practices in questionnaire design and privacy, definitions of sexual contact and sexual assault, and interviewer effects.

Respondent Reactions to a Sexual Victimization Survey in a Multi-mode Pilot Study

Suzanne Kaasa, *Westat*
Darby Steiger, *Westat*
Leanne Heaton, *Westat*
David Cantor, *Westat*
Rose McAloon, *Westat*

Sexual victimization surveys are often faced with scrutiny from Institutional Review Boards (IRBs), service providers, and the public because of their sensitive nature (e.g., Jaffe et al., 2015; Baldor, 2013). Prior research demonstrates that sensitive surveys produce more self-reported benefit (e.g., increased self-awareness, empowerment) than harm (e.g., distress and unwanted thoughts), and that most respondents who experience negative feelings do not regret participation (e.g., McClinton, 2015; Newman et al., 2006). However, few studies have examined how reactions may be affected by different survey data collection methods such as interviews conducted via phone versus in person. In addition, researchers often offer resources to respondents to address potential distress, but rarely measure the

impact of these efforts. We address these gaps by presenting data collected from over 10,000 women who participated in a pilot study to test alternative ways of collecting data on rape and sexual assault for the National Crime Victimization Survey (NCVS) on behalf of the Bureau of Justice Statistics. Data were collected by telephone with an interviewer using a Computer Assisted Telephone Interview (CATI) or as a self-administered questionnaire using Audio Computer Self-Administered Interviewing (ACASI). Samples for each type of interview were recruited using probability and non-probability methods. At the conclusion of each interview, a series of debriefing questions were asked to assess reactions to the survey experience and sexual assault service resources were provided. Differences in levels of self-reported emotional reactions to the survey will be compared across data collection modes, and sample types. In addition, we will present data on respondent utilization and perceived usefulness of resource information provided by researchers. Implications of these findings for conducting future research and communicating with respondents, IRBs, service providers, and the public will be discussed.

Advancements in Cross-cultural and Multilingual Questionnaire Design and Pretesting

Panel organizer: Mandy Sha, *RTI International*

The proposed panel includes five papers that present the leading methods for conducting multilingual research and substantive issues surrounding US Census coverage among non-English speakers. This panel is of interest to methodologists and practitioners who would like to improve measurement comparability across languages, design linguistically and culturally appropriate instruments, and encourage participation among non-English speakers. The panel is chaired by Dr. Patricia Goerman who heads the Language and Cross-Cultural Research Group at Census Bureau's Center for Survey Measurement. Dr. Ana Villar, an expert in cross-cultural research, serves as the discussant. In preparation for the 2020 Census, the Census Bureau developed a comprehensive plan of qualitative research designed to pretest the 2020 Census questionnaires and the materials used to encourage participation in the Census. The research was conducted using English, Spanish, Chinese, Korean, Vietnamese, Russian, and Arabic languages and in three data collection modes: via the Internet, by in-person enumerator visits, and by paper self-administered format. A total of 384 respondents participated in cognitive and usability interviews and around 300 participated in focus groups. We propose to start the panel by providing an overview of the methods and the results of this multilingual and multimode research study. Next, we discuss optimizing the visual design and usability of government information to facilitate their access by Asian non-English speakers. In addition, we identify unique issues regarding symbols and layout in Arabic language questionnaires and pinpoint the challenges these issues will present to survey data if unaddressed. We also examine whether immigration waves among Russian speakers (e.g. Soviet era vs. recent immigrants) affect their survey responses, as well as how to recruit and interview this population. Finally, we conclude the panel by evaluating the appropriateness of utilizing vignettes in cross-cultural survey research, which compares data from focus groups and cognitive interviews across 7 languages.

Multilingual Language Research at the U.S. Census Bureau in Preparation for the 2020 Census: Looking Across Modes and Languages

Patricia Goerman, *U.S. Census Bureau*
Hyunjoo Park, *RTI International*
Katherine Kenward, *Research Support Services*

The U.S. Census Bureau is committed to increasing research and pretesting related to non-English language materials for use in our censuses and surveys. As such, a great deal of work is being done related to pretesting non-English forms and other materials in preparation for the 2020 Census. In particular, in 2015 we undertook a large study to test Spanish and corresponding English-language materials in three modes: 1) the interviewer-administered, CAPI, mode, 2) paper questionnaires and 3) the internet mode. In addition we have tested self-administered questionnaires for use in Chinese, Vietnamese, Korean, Russian and Arabic. Finally, we have conducted focus groups in all seven languages in order to look at translation issues and the development of verbal and written messages for use with speakers of these languages. This paper will provide a high level summary of the types of findings that have resulted from the various pretesting methods described above. Testing in Spanish and English across the three different modes has shown us that some of the same cognitive issues appear across modes but there are also mode specific issues that arise. Looking across languages, we have seen that different questions and formats can cause difficulty for different populations. We will discuss next steps in terms of language testing at the Census Bureau.

The Hidden Barriers: Assessing Usability of Government Websites for Asian Non-English Speakers

Y. Patrick Hsieh, *RTI International*
Hyunjoo Park, *RTI International*
Mandy Sha, *RTI International*
Patricia Goerman, *U.S. Census Bureau*

Usability and accessibility accommodation for people with disabilities, guided by Section 508 of the Rehabilitation Act, have been important benchmarks for evaluating the effectiveness of e-government, which refers to the utilization of information and communication technologies for delivering government information and services. While low English proficiency is not a physical disability, non-English speakers face usability and accessibility issues when they do not understand the English written text that is part of the visual design of websites. A common solution is providing translations to facilitate non-English speakers' access. However, there is scant research on the usability of translated materials presented on government websites, and on how non-English speakers interpret materials and visual cues whether or not translation is provided. In this study, we seek to explore usability barriers stemming from language barriers. We observed how 45 Chinese, Korean, and Vietnamese speakers, who spoke little or no English, interacted with possible government website designs that were translated into their native languages. An in-depth debriefing followed the usability/cognitive interview. The data came from a US Census Bureau study that pretested the translation of the 2020 US Census form and the design and content of a potential hosting website for the multilingual Census forms. Our results show that while participants may recognize the interactive website features (such as pull-down menus) in spite of the presence of English labels, the language barrier may discourage them from engaging further with the website. Many users also explored and clicked on random English texts for potentially embedded hyperlinks which resulted in unsuccessful attempts at locating and accessing the intended URLs due to their limited English proficiency. We

conclude by discussing the implications of our study for optimizing the visual design and usability of translated websites to maximize access to government information for non-English speakers.

Symbols and Layout: Unique Issues in Arabic Translation of Self-administered Survey Forms

Mandy Sha, *RTI International*
Sama Harp, *RTI International*
Mikelyn Meyers, *U.S. Census Bureau*

Written language is a semiotic system of representation, consisting of words and signs (symbols, punctuation) (Halliday and Hasan, 1985). Not all languages and cultures share the same system or interpretation of symbols and signs. The difference between English and non-English writing practice creates challenges in the translation of self-administered survey forms as well as form navigation. Prior studies investigating symbols and layout in survey translation were based in the Chinese language (Pan et al, 2005; Sha and Pan 2011) and there is no literature that provides comprehensive documentation or findings from additional in-language cognitive interviews. This paper aims to investigate and identify issues unique to the Arabic language and practice and pinpoint the challenges these issues will present to survey data if unaddressed. This study uses data from 30 Arabic language cognitive interviews from a project conducted by the U.S. Census Bureau. This project pretested the 2020 US Census form in the paper format. We found that Arabic language rules and writing practice affected form navigation. The symbol "X" that is commonly used to denote "yes" in a response means "no" according to Arabic writing practice. Because Arabic script must be written in cursive form and cannot be in disconnected letters, it is not possible to "translate" the instructional function of the ALL CAPS text. In addition, Arabic naming convention and right to left writing orientation directly affected form layout and design to collecting accurate name, address, and phone numbers. We designed some solutions to mitigate the effect of these issues and will discuss the feasibility of our recommendations. This paper will be of interest to questionnaire designers who provide language assistance to respondents with limited English proficiency.

Qualitative Research with Demographically Diverse Russian Speakers

Alisú Schoua-Glusberg, *Research Support Services*
Katherine Kenward, *Research Support Services*
Gerson Morales, *U.S. Census Bureau*

One of the top languages spoken in the U.S. in linguistically isolated households is Russian. Speakers are immigrants from diverse national origins, as Russian speakers still abound in most of the Former Soviet Union countries. Since 2010, the U.S. Census Bureau has provided a Russian language questionnaire. As part of the multilingual research for the 2020 Census, we conducted 30 cognitive interviews over two rounds and four focus groups with Russian speakers to qualitatively test the census form and other census materials with the target population. The focus of interest is the translation of the materials to make sure that the terminology used is interpreted as intended, and that speakers who have come to the US from a variety of countries and realities can understand the language well. This paper will cover issues and successful strategies in recruiting this population, as well as challenges and obstacles. Furthermore, we will examine to what extent coming of age in the Soviet era vs. younger immigrants affects how research participants interpret questions and view respondent materials. We will present data from the interviews and focus groups to illustrate our findings.

The Performance of Vignettes in Focus Groups and Cognitive Interviews in a Cross-cultural Context

Mikelyn Meyers, *U.S. Census Bureau*
Yazmin Garcia Trejo, *U.S. Census Bureau*
Lucia Lykke, *U.S. Census Bureau*
Nicole Holliday, *U.S. Census Bureau*

Survey researchers conducting pretesting via cognitive interviews or focus groups often utilize vignettes to evaluate questions and answer categories in order to identify possible measurement problems, particularly for reporting situations that are relatively rare or for sensitive topics (Willis 2015; Martin 2004). In a cross-cultural context, vignettes can also be used to help anchor responses on a common scale of measurement (King et al 2004). While research has been done regarding the comparative performance of vignettes in certain languages (Goerman & Clifton 2011, Sha & Pan 2009), there is relatively little research that compares the performance of vignettes across pretesting methodologies in languages other than English. To address this gap, in this study a vignette that tested a problematic homeownership status question was administered in focus groups and cognitive interviews conducted in seven languages: English, Spanish, Chinese, Korean, Vietnamese, Russian, and Arabic. The resulting interactions were coded to quantify the responses given to the vignette, as well as the types of problems uncovered with the question or answer categories. In order to account for the tendency that respondents might have to change their answer to the vignette once they hear how others responded in the focus group, we collected a record of respondents' written answers to the vignette before the focus group discussion began. We examine how the richness of the data from administering a vignette in focus groups compares to the data from cognitive interviews in each language. Additional analysis uncovers discrepancies between written responses and focus groups discussion, with trends observable by language. Preliminary analysis shows that while many respondents had problems discerning the focal point of the question (e.g. the housing unit or the respondent), the problems were more pronounced in certain languages and methodologies than in others.

Blazing a Trail Without Losing the Path: How the History of LGBTQ Public Opinion Research Can Lead Emerging Issues and Questions

Panel organizer: Gretchen McHenry, *RTI International*

Recent AAPOR conferences have had excellent papers presented on both the history and development of LGBTQ focused research as well as what researchers are currently doing to handle the complex measurement issues around identity and attraction. This year, we will go a step farther by pulling together a panel of diverse researchers to talk about what the next steps might be for work on LGBTQ issues and how we get there. Our panelists will discuss the growth of research on LGBTQ issues over time, what are the pressing current questions, what issues we will see arise in the future, and how their own work has impacted and will continue to spearhead the growth of LGBTQ research. The panel will consider the following questions: - Where have we been and where are we going? - Have methodological advances led to asking new questions? - Are there any enduring questions? - What are the things that we ought to be looking at? - Are there methodological obstacles to addressing these questions? - What are the best ways to access and sample the LGBTQ population? Moderated by Dr. Kenneth Sherrill, Professor Emeritus, Political Science Department at Hunter College

Reflections on Over a Quarter Century of Using Surveys to Study Sexuality and LGBT Populations
Stuart Michaels, *NORC at the University of Chicago*

In 1988 I began work on what would see the light of day as the 1992 National Health and Social Life Survey (NHLS) which was the first comprehensive probability based survey of sexual practices of the adult (18-59 year old) population in the U.S. While involved in all aspect of the design and execution of the NHSLS, my primary research interest was on the materials relating to homosexuality. The NHSLS included questions on sexual partners, practices, desire, and self-identification. A major concern driven by HIV/AIDS at the time was on population estimates of prevalence but the research also emphasized the complex inter-relationship between behavior, attraction, and identity. This continues to be a major concern of my research. How do we conceptualize and measure a multi-dimensional construct such as sexual orientation? Major progress has been made in the number, size, and quality of surveys that include measures related to sexuality and sexual identity, but more is needed. More recently I have been involved in research using Respondent Driven Sampling to study the social networks and sexuality of young Black men who have sex with men. This research along with the progress of the transgender population and movement have pushed us to extend our work to develop and include measures of gender identity and expression. An ongoing challenge is to generate large enough samples of LGBT persons that are as representative as possible to study the complexities of gender, sexuality, and the social factors such as networks, social context, and stigma that affect the well-being of LGBT populations broadly defined. To accomplish this we need to continue to improve our theory, measures, and sampling strategies.

Intersectionality and Engagement among the LGBTQ+ Community
Samara Klar, *University of Arizona*

As the lesbian, gay, bisexual, transgender, and queer (LGBTQ+) community makes progress toward attaining equal rights, a growing body of scholarly attention is focusing on this increasingly visible minority group. Yet, studies of attitudes among LGBTQ+ Americans themselves remain limited due to small sample sizes and scarce data. As a result, scholarly work on LGBTQ+ issues is almost entirely devoted to measuring straight America's opinions. In this study, we administer both a survey and an experiment to a large and diverse sample of LGBTQ+ Americans. Our findings are two-fold. First, we demonstrate that intersectionality has important effects on attitudes within the LGBTQ+ community. Specifically, LGBTQ+ respondents who are at the intersection of multiple minority groups display dramatically lower levels of political engagement. Second, we test the mobilizing influence of out-group versus in-group cues on LGBTQ+ Americans. In line with previous work, we find that government action to support a threatening out-group engages LGBTQ+ Americans to participate in politics, whereas government action to support their own in-group has a significantly smaller effect. These findings help us to understand an increasingly politically active subset of the electorate and, more broadly, shed light on the influence of intersectionality on opinion and political engagement.

Examining Public Support for Transgender Rights and Supportive Policies
Ashley Kirzinger, *University of Illinois at Springfield*

Dr. Jason Pierceson and Dr. Ashley Kirzinger are currently collaborating on research examining both attitudes towards and knowledge of transgender inclusive policies in the U.S. Recently, a great deal of attention has been paid to lesbian and gay rights, particularly the issue of marriage equality. While

the transgender rights movement is often conflated with the lesbian and gay rights movement, transgender individuals face unique challenges that require a different policy response in many circumstances. In addition, researchers in this area face their own unique challenges as there is not much public opinion data exists gauging levels of support for transgender-supportive policies and very little is known on how much individuals know about issues surrounding transgender persons including awareness of specific terms and policies. The fact that the majority of Americans do not know much about these policies poses a problem for researchers as "ask[ing]people a question in a poll they haven't really thought about…will get a useless answer" (a word of wisdom from the late Andy Kohut cited by Tom Rosentiel's "Andrew Kohut: An appreciation, Poynter, September 8, 2015). In order to assess both attitudes and level of knowledge on these issues, Drs. Pierceson and Kirzinger developed an online survey using Qualtrics market research panels. While online surveys face issues of probability sampling, they allow researchers to collect significant paradata including question response time. Drs. Pierceson and Kirzinger strongly believe that using online methodologies to examine the overall responses in conjunction with the paradata provides key insights into the firmness of attitudes as well as the successfulness of specific question wording. We are interested in sharing the results from our most recent survey as well as directions for future research in this area.

New Opportunities for Sampling LGBT Populations
Ilan H. Meyer, *The Williams Institute*

Research on LGBT population has relied mostly on non-probability samples. Recently, probability samples became more routinely available to researchers of LGBT populations as a growing number of national and local surveys add questions about sexual orientation and gender identity (SOGI) to the list of population characteristics they inquire about. But challenges for researchers abound: First, primarily due to the low prevalence of LGBT identity in the general population, even large surveys provide only a small number of LGB respondents. This often precludes analysis by race/ethnicity, age groups, and other socio-demographic variables of interest. Second, even if sample sizes were sufficient for analyses of interest, the survey items available in general population surveys are not typically selected with the LGBT research questions in mind specifically. Thus, researchers interested in special content relevant to LGBT populations are hindered when using available surveys. To address these challenge, I introduce a new methodology. In two new studies, researchers are using a 2-step recruitment method whereby the general population is first screened with the SOGI question and those identified as LGBT are then invited to participate in the targeted research. The full interview is administered only to the LGBT individuals. Potential costs of this approach are high, depending on the desired sample size due to the need for a very large first-step screen. Costs can be curtailed by collaborations with organizations already conducing research that involves the screen of large populations. I will describe using this method in two new NIH-funded studies where investigators plan to collect a sample of close to 2,000 LGBT individuals for the study of stress, identity, and health of LGBT individuals in the U.S.

Using Qualitative Methods to Improve Data Quality

Context Effects in Anchoring Vignette Questions

Mengyao Hu, *University of Michigan*
Sunghee Lee, *University of Michigan*

One major source of measurement error in self-assessments is reporting heterogeneity that arises due to differences in respondents' usage of response scales. Anchoring vignette is a method developed to correct for reporting heterogeneity in interpersonal and cross-cultural comparisons, which have been used to evaluate survey reporting behaviors in studies on various topics (e.g., health, political efficacy and life satisfaction) and included in a wide range of surveys, such as the World Health Surveys (WHS), the Study on Global Ageing and Adult Health (SAGE), the Health Retirement Survey (HRS) and the Survey of Health, Ageing and Retirement in Europe (SHARE). Although widely used, few studies have evaluated the context effects of anchoring vignette items, such as the effects of different response scales on anchoring vignette method, the selection and ordering of vignette items corresponding to the self-assessment questions. Such context effects, if present, can potentially influence the method's controlling effect of reporting heterogeneity. To evaluate the potential context effects with anchoring vignettes, a randomized online survey experiment is conducted through Amazon Mechanical Turk. In this experiment, we randomized the order of self-assessment questions relative to vignette questions, used different response scales of the questions – five-point scale and 100-point scale, and included differently designed vignettes for the same domain. This study will evaluate whether responses to the vignette questions are subject to survey context effects, and how the application of anchoring vignette methods can be potentially influenced. The result of this study will inform and sensitize researchers about the potential sources of measurement errors in anchoring vignette items, and provide suggestions to better design vignette items.

Using Web Panels to Quantify the Qualitative: The National Center for Health Statistics Research and Development Survey

Paul J. Scanlon, *National Center for Health Statistics*

As with any qualitative method, one of the major limitations of cognitive interviewing is the fact that its findings cannot be generalized from its purposive sample to the larger, statistical sample of a survey. In 2014 the National Center for Health Statistics launched the Research and Development Survey (RANDS), a web-mode survey using the Gallup Panel whose purpose was to explore how web panel surveys may be used to supplement official health statistics and the methodological work at NCHS. One of the main research questions RANDS was developed to explore was whether or not web panels were appropriate vehicles for attempting to extrapolate qualitative findings from question evaluation studies through the use of "embedded probes." Embedded probes are structured cognitive probe questions that are administered directly following a survey item that is under evaluation. Unlike typical cognitive question evaluation methods, these probes are administered as part of a fielded survey, and not in a pre-test or lab environment. In this research, National Health Interview Survey (NHIS) questions were evaluated using in-depth, face-to-face cognitive interviews. Following the analysis of this data, patterns of interpretation that respondents used to answer the NHIS questions were analyzed and the resulting cognitive schema were developed. Using these schema, structured probe questions were designed and then administered following their corresponding NHIS questions in the RANDS. Here, we present the initial findings of the RANDS analysis, showing that the use of embedded probes on a web panel survey are a relatively cheap and efficient way to extrapolate the results of qualitative question evaluation

studies to a wider survey population, and to show the different patterns of interpretation of questions across various groups of respondents.

Practical Considerations for Using Vignettes to Evaluate Survey Items

Darby Steiger, *Westat*
Douglas Williams, *Westat*
W. Sherman Edwards, *Westat*
David Cantor, *Westat*
Jennifer Truman, *U.S. Bureau of Justice Statistics*

Vignettes are a common tool to assess how people think about specific concepts, allowing data to be gathered on survey questions that might not normally apply to the respondent (Martin, 2004). They are also useful as "a less threatening way to explore sensitive subjects" (Finch, 1987). To test potential new screening items for the National Crime Victimization Survey (NCVS), Westat presented seven vignettes about incidents of personal victimization to a non-probability web panel. Respondents were asked how they would answer a NCVS screening question based on each vignette. Each vignette had four to six versions in a factorial design, with vignette versions independently randomized for each respondent. Factors varied the seriousness of the crime, who committed the crime, and other conditions. Respondents were assigned to report how they would respond to current NCVS item wording, vs. a "streamlined" version which included fewer examples of the victimization type. Surprising results were found, with high levels of respondents indicating that they would not consider the vignettes to be a crime, even using the current NCVS wording. To understand why reporting results were lower than expected, we conducted a series of 18 follow-up cognitive interviews. Our hypotheses were 1) that the number of examples and cues in the NCVS items might be leading to cognitive overload; and 2) that respondents may be focusing on a salient characteristic of the vignette which may be mitigating the interpretation of it as a victimization. Cognitive interviews revealed few problems with either of the versions of the NCVS questions. Rather, participants revealed it was the vignettes themselves that caused difficulty. This paper will discuss findings from the survey and cognitive interviews, and will present practical considerations for using vignettes in survey research.

How Long Did That Take?: Understanding Differences Between Diary and Stylized Measures of Time Use

Robin L. Kaplan, *U.S. Bureau of Labor Statistics*
Brandon Kopp, *U.S. Bureau of Labor Statistics*
Polly Phipps, *U.S. Bureau of Labor Statistics*

Data on how people spend their time is of great interest to researchers to better understand economic and societal trends. One common data collection method is time diaries, where respondents report on their activities for a specified period of time. For example, in the American Time Use Survey (ATUS), interviewers use a set of scripted open-ended questions to walk respondents chronologically through their activities during the prior 24-hour day. In contrast, stylized questions ask people about "the average, normal, or typical" amount of time spent doing activities. But diary and stylized measures often yield different time use estimates for activities as varied as work, exercise, and sleep. For example, the ATUS reports that respondents sleep approximately 8.6 hours per night, whereas surveys using stylized measures report approximately 6.9 hours. In this research, we build off previous qualitative work showing that differences between diary and stylized were tied to factors at each stage of the response process (e.g., comprehension of the term "sleep" and social desirability concerns). We

recruited participants from online panels (N=1200, Study 1) and 400 participants using a quota sample designed to represent demographic groups (Study 2) to assess differences across diary and stylized measures. We randomly assigned participants to one of three survey topics: employment, health, or a control group, and whether participants received definitions of activity terms. Participants completed an activity log where they reported all of their activities during the previous 24-hours, along with a subset of stylized questions about their activities, in counterbalanced order. We examine differences in reported time spent on activities across diary and stylized measures, and factors that may lead to systematic differences in reporting across both measures (e.g., survey context and providing definitions of activities). We discuss implications of these results for understanding diary versus stylized measures of time use.

May 14, 2016
Poster Session #3

The Benefits of an Increased Cellphone Allocation for Dual-frame Surveys to Target Low Socioeconomic Persons

Jamie Ridenhour, *RTI International*
Marcus E. Berzofsky, *RTI International*
Bo Lu, *The Ohio State University*
Caroline Blanton, *RTI International*
Tim Sahr, *Ohio Colleges of Medicine Government Resource Center*

Research has shown that the rate of movement to majority cellphone use from partial or complete landline use is not proportional throughout subgroups within the U.S. population (see, e.g., Lu, et al., 2014). These studies have found that young persons with children, minorities, and those with income near the Federal Poverty Level have made the shift more quickly than other demographic groups. Therefore, dual-frame telephone studies such as the Ohio Medicaid Assessment Survey (OMAS), which are interested in examining low socioeconomic status (SES) subpopulations, need to increase the allocation of their sample that come from a cellphone frame. Moreover, even though data collection for cellphone samples generally costs more than landline samples, the increased rate of interviewed sample members from key subpopulations mostly negates the cost-per-complete difference. In 2012, 25% of the OMAS sample was randomly chosen from the cellphone frame. For 2015, the sample allocation was increased to 50%. This paper presents the impact of the increased OMAS sample allocation to the cellphone frame on respondent yield and data collection costs and serves as a guide for similar dual-frame surveys.

Recycling in New York City

Janet L. Streicher, *Baruch College, CUNY*
Micheline Blum, *Baruch College, CUNY*
Doug Muzzio, *Baruch College, CUNY*

The New York City Department of Sanitation and Baruch College Survey Research conducted a series of surveys focused on enhancing understanding about New Yorkers attitudes, behaviors, and level of engagement related to recycling and waste. The research was designed to inform DSNY recycling programs and communications and message development. Findings and insights from these studies are also germane to other large city initiatives that include economically and culturally diverse populations

and neighborhoods. Two of the recycling surveys will form the basis of the presentation 1. Organics Collection Pilot Area survey (n=500). This is a RDD sample within the specified pilot areas. 2.General Population survey of NYC household (n=2000+ including 400 from zip codes that have low incidence of recycling). This is an RDD survey with dual frame for low in 3.A third survey, which may not be completed prior to the conference, will be among NYC Public Housing households. All respondents are surveyed using computer-assisted telephone interviewing (CATI), and are given the option of being interviewed in English, or Spanish. Mandarin was also available (NYC has less than 1% Mandarin speaking households). Key findings will be presented: -Differences in thought and action relating to recycling. -Motivations related to recycling. -Resistances to recycling -Comparisons among different NYC neighborhoods, including Public housing. - Influence of neighbors on recycling behavior -Attitudes toward recycling different types of waste Thank you, Micheline Blum, Doug Muzzio, Janet Streicher

Contact History Instrument: Using Data from Field Staff to Inform Data Collection Efforts

Caitlin Waickman, *NYC Department of Housing Preservation and Development*
Lin Wang, *NYC Department of Housing Preservation and Development*
Tristan Brennan-Torell, *NYC Department of Housing Preservation and Development*
Daniel Goldstein, *NYC Department of Housing Preservation and Development*

In recent years, with the rising cost per complete in survey research, it is increasingly important to maximize response rates by using tools and programs that will further field efforts without putting a financial burden on the study. The New York City Housing and Neighborhood Study (HANS), a collaboration between the NYC Department of Housing Preservation and Development and Columbia University, Teachers College, currently employs instruments and tools developed by research staff to cost-effectively enable strategic data collection. Modeled on the U.S. Census Bureau's innovations, the project uses a Contact History Instrument (CHI) and case dashboards. These tools have been adapted to a more resource-limited environment and provide a model for how other projects may implement and utilize similar tools with non-specialized, widely available programs. The CHI is a short survey that Field Interviewers complete after every contact attempt to record information on the attempt, strategies employed, and new contact information obtained. After extensive transformation using widely available statistical programs, Excel formulas populate these dashboards. The dashboards, saved as a single PDF, allow field interviewers to approach each case as a collective body, with a full understanding of past attempts and best times to reach potential respondents presented both textually and graphically. The HANS CHI and dashboards enable the field staff and research team to track and approach cases individually, and in detail, using programs that can be updated daily but do not put a financial strain on the study. This poster describes how the CHI and dashboards are created and how they are utilized daily by HANS field staff, providing guidance to other projects on how to employ these sophisticated tools using simple means.

Why Aren't You on Social Media? … It's About Time

Ann E. Williams, *Georgia State University*

This paper employs longitudinal data from the General Social Survey (GSS) to explore the unique characteristics of individuals who are not active on social media platforms – focusing specifically on social media non-adoption. Drawing from literature in media studies and economics, the paper tests a series of "perceived time-displacement" hypotheses that explain significant portions of individuals' reluctance to adopt social media. The analyses link non-adoption to three unique forms of time

perceptions – the first is economically-oriented, the second is value-oriented, and the third is affectively-oriented. All three forms of time perception correlate with non-adoption decisions in different but significant ways, with economic-orientations playing the greatest role. In addition to documenting the unique factors that contribute to the "lost time" phenomenon, the analyses also indicate that individual predispositions – including age, gender, and household composition – play significant roles in explaining social media non-adoption. The relationships manifest in interesting ways over-time, with results indicating that young people and those with large household networks increasingly constitute the largest share of non-adopters.

"Screen and Go": Using New Technology to Facilitate a Fluid Transition Between a Household Screener and CAPI Interview in a Population Survey

Anna F. Wiencrot, *NORC at the University of Chicago*
Heather Leaver-Spear, *NORC at the University of Chicago*
Daniel Lawrence, *NORC at the University of Chicago*
Reshel Kurniadi, *NORC at the University of Chicago*
Shelly Plummer, *NORC at the University of Chicago*
Katie O'Doherty, *NORC at the University of Chicago*

Screening households to determine eligibility for a study is a longstanding practice in survey research. Managing the transition from identifying an eligible respondent and interviewing that respondent often necessitates multiple visits, multiple devices, or other logistical challenges. For the third wave of the National Social Life, Health and Aging Project (NSHAP), NORC developed an innovative approach to household screening which provided a seamless transition from screener to interview on a single device. We called this initiative "Screen and Go." The ability to transition from screener to interview on a single device allows the field interviewer to capitalize on the cooperation gained during a successful administration of the household screener. The NSHAP interview included a complex computer-assisted personal interview (CAPI) requiring a full-size laptop while the device used for screening needed to be easy to use while standing at the front door. We selected a dual-mode laptop that allowed the household screener to be conducted in tablet mode, but supported the full NSHAP CAPI in laptop mode. If an interviewer identified an eligible respondent through the household screener, that respondent case was created in our case management system and the interviewer would be able to switch from tablet mode into laptop mode to launch the NSHAP questionnaire immediately. The development and fielding of this new technology proved both challenging and successful. This paper examines the Screen and Go initiative from ideation to field use. Our findings will help other studies interested in closing the gap between screening and interviewing or requiring both touch-screen tablet and laptop functionality in a single device.

Measuring Stress: Adequacy of a Single Scaled Response - A Comparison of a Single 10-point Measure vs. Cohen's Perceived Stress Score

Michele Salomon, *Nielsen*
Aimee Vella-Ripley, *Nielsen*

Since 2007, the American Psychological Association (APA) has commissioned Harris Poll to conduct The Stress in America™ survey measuring attitudes and perceptions of stress among the general public and identifies leading sources of stress, common behaviors used to manage stress and the impact of stress on our lives. The results of the survey draw attention to the serious physical and emotional implications of stress and the inextricable link between the mind and body. Each year prior thirty day

stress is measured using a simple 10-point scale where 1= little to no stress and 10=a great deal of stress. Over the course of the past three years (2013, 2014, 2015) we also employed Cohen's Perceived Stress Scale (PSS), a self-reported instrument with 10 items which has been widely used as a measure of psychological stress where respondents self-report the degree to which situations in their life are appraised as stressful over the prior thirty day period. Our analysis focused on measuring the reliability of the simple 10-point scale in accurately capturing adults' stress levels and to determine how correlated these results were compared to the full 10-item PSS instrument (used in 2013 and 2014) and the shorter 4-item version (used in 2015). Our analysis compared the scores from the PSS with the single 10-point measure and found a very high correlation in response patterns. Our findings identified that regardless of measure, similar proportions of adults fell into a highly stressed category. Findings from the APA research are consistent with results of the PSS research in finding that women report more stress than men and that older adults report less stress than younger adults. Our research strongly suggests that the single measure adequately captures adults understanding of their stress compared to the detailed PSS and can be used as a reasonable substitute.

Interviewer Gender: The Topics and Constructs that are Most Sensitive to Gender Differences in Interviewers and Respondents in CATI Surveys
Stephanie Marken, *The Gallup Organization*
Sofia Kluch, *The Gallup Organization*

Despite considerable research on how male and female respondents may differ on a variety of public opinion topics, including governance and the economy, there has been limited examination of how the gender match between the interviewer and respondent might impact respondents' answers on different question types for CATI studies. This study utilizes Gallup World Poll data from Italy, Portugal, the United Kingdom, and Norway and Gallup Daily tracking data from the United States to examine four different gender match categories; female interviewer- female respondent, male interviewer-male respondent, female interviewer-male respondent, and male interviewer-female respondent. The four groups are compared on items related to safety, civic engagement, economics, governance, and wellbeing – along with a focus on differences in demographic data, specifically age, marital status, and education. Early findings suggest consistent distinctions across populations for same-sex interviewer-respondent pairs, potentially suggesting that respondents are more (or less) honest with a member of their own sex during an interview. The analyses is divided into items or topics where male and female respondents differ from each other with a female interviewer (many distinctions); topics where male and female respondents differ from each other with a male interviewer (few distinctions); topics where same-sex interviewer-respondent pairs differ from mixed-sex interviewer-respondent pairs (many distinctions); and finally where male interviewers-respondent pairs differ from female interviewer-respondent pairs. Many of these differences are not evident at the overall respondent gender level, but manifest themselves when examined at the interviewer gender level. Among the many items being examined there is one World Poll item that specifically asks if women in the country are treated with respect and dignity. Despite expectations, this item did not yield significant differences in any country based on interviewer-respondent match.

Using the Data Documentation Initiative to Document the Consumer Expenditure Survey

Daniel W. Gillman, *U.S. Bureau of Labor Statistics*
Reginald Noel, *U.S. Bureau of Labor Statistics*

The Data Documentation Initiative (DDI) is a set of statistical metadata standards for documenting social science data. Currently, there are two series, Codebook - for documenting data sets one at a time; and Lifecycle - for documenting the data sets, questionnaires, and the survey lifecycle for any number of surveys. Each has a number of versions, all of which are in use. Social Science data libraries, data archives, national statistical offices, and international organizations around the world are all making use of the standard. The US Bureau of Labor Statistics (BLS) chose DDI-Lifecycle version 3.2 and the commercial DDI software Colectica Designer to build a pilot system for documenting the Consumer Expenditure Survey (CE). This work was inspired by the realization that publishing PDF files for documenting each year of the CE public use microdata files (PUMD) was inefficient. For instance, using the yearly files, it would very hard to determine whether a variable is used unchanged over every year the PUMD files are published. The DDI-Lifecycle promises a solution to this problem and many others like it. In addition, the DDI-Lifecycle is designed to document the survey lifecycle. This means the design, processing steps, and subsequent data transformations can be described. As CE is undergoing a redesign at this time, the opportunity presents itself for implementing a metadata approach to documentation. BLS has a pilot project to determine the effectiveness of DDI-Lifecycle and the Colectica Designer software. In this poster, we will describe the advantages, accomplishments, pitfalls, and lessons learned in using the approach DDI. We hope to attract other survey groups to try the same approach.

Using Paradata from a Pilot Study to Develop Strategies for Monitoring Response in a Longitudinal Component of the National Health and Nutrition Examination Survey

Debra J. Brody, *National Center for Health Statistics*
Vicki L. Burt, *National Center for Health Statistics*

Analyzing paradata has become an important focus to understand the data collection process and its role in developing strategies for improving survey response. While this auxiliary data has been useful for many other surveys and organizations, few studies have examined the use of paradata in the National Health and Nutrition and Examination Survey (NHANES), which uniquely requests participation by in-person interview and physical examination. In a recent NHANES follow-up pilot study, 223 adult participants were asked to collect an oral mouthwash rinse sample at home, six months after an initial baseline collection in the NHANES mobile examination center. Participants were also asked to complete a short set of questions, on the Internet, telephone, or mail. Extensive follow-up (text, telephone, mail) of non-respondents was conducted, beginning one week after the home collection oral mouthwash rinse kits were mailed, up until six weeks, following. The purpose of the pilot was to evaluate the response rates for both the oral mouthwash sample collection and the questionnaire, and to determine the quality of a home collection oral mouthwash rinse sample, in planning for a NHANES longitudinal study on the natural history of the oral human Papilloma virus infection. In our analysis, we examine survey participant level-data and paradata from the baseline NHANES interview and examination together with paradata from our pilot study (contact information, mode of contact, intensity of location effort, time to respond) to understand the factors related to participation in the follow-up component. Findings will be used to create new and automated methods to extract paradata during the data collection process, as well as to monitor response rates in the future longitudinal study.

Try, Try, Try Again: The Iterative Development of NCSES's Microbusiness Survey
Jennifer Crafts, *Westat*
Audrey Kindlon, *National Science Foundation, National Center for Science and Engineering Statistics*
Brad Chaney, *Westat*

The National Center for Science and Engineering Statistics (NCSES) is developing a new survey, the Microbusiness Innovation Science and Technology (MIST) survey, in response to the National Academy of Science's Committee on National Statistics (CNSTAT) recommendation to explore ways to measure innovation and research and development (R&D) activities in microbusinesses (companies with fewer than 5 employees). Due to the "micro" nature of the target business population, NCSES expected the survey to have elements in common with a household survey. Therefore, NCSES incorporated into the design and development process a combination of qualitative methods used for household surveys as well as establishment surveys. The overall survey development approach included exploratory interviews, cognitive interviews, usability tests, and debriefing interviews with microbusiness owners. Westat assisted NCSES in evaluating feasibility and motivation to respond, perceptions of the response process, question comprehension, and need for definition of concepts such as innovation and R&D. Development began with iterative rounds of exploratory interviews which focused on relevance of questions that have been asked of larger businesses; and if these topics/questions were relevant, how best to tailor them for microbusinesses. Feasibility, perceived burden, and motivation to respond were also assessed. Iterative rounds of cognitive interviews focused mainly on question comprehension and response process (steps respondents would need to take to provide the data), while continuing to assess relevance, feasibility, and motivation. Usability testing assessed issues in providing response via the web-based version of the survey. Debriefing interviews focused on response issues that surfaced in the pretest and the pilot test. This presentation will highlight for survey designers the importance of employing a variety of qualitative methods when developing a new survey for an extremely diverse survey population, and also the need to iteratively revise the content based on respondent motivation and feasibility of providing the requested data.

Effectiveness of the Think-aloud Method in Children of Different Age Groups
Mila Sugovic, *EurekaFacts, LLC*
Michael Plotkin, *EurekaFacts, LLC*
Bohdana Sherehiy, *EurekaFacts, LLC*

This study address the need to apply different cognitive interviewing techniques, such as the think-aloud method, when conducting survey-item and usability testing with children. Research shows that young children typically find the think-aloud method to be difficult (Someren et al, 1994), however, it is unclear whether this difficulty stems from a lack of adequate verbalization skills, the general content of their thought process or difficulty related to concentrating on problem-solving. Importantly, there is little empirical research on the effectiveness of the think-aloud method across children of different age groups. This study examines the application of a think-aloud technique during testing of a computer-based assessment, across two different grades and age-groups: grade 4 (age 9-10) and grade 8 (age 14-15). We compared the effectiveness of the think-aloud method between the two grades by evaluating the verbal reports length, time of response, and the number of reported problems across items. We found that the think-aloud method varies in the quality of the verbal output obtained from children in the two age groups. Thus, different cognitive interviewing strategies are necessary for obtaining

verbalization data from children of different ages. We also discuss the parallels between children's ability to verbalize thoughts at different ages and psychosocial development stages. Specifically, how the knowledge on different developmental stages may affect the cognitive interviewing techniques adaptation for children of different age groups.

Small Area Estimations for Health Policy Survey Research: A Comparison of Direct, Synthetic and Composite Estimators using the 2015 Ohio Medicaid Assessment Survey

Daniel Weston, *Ohio Colleges of Medicine Government Resource Center*
Tim Sahr, *Ohio Colleges of Medicine Government Resource Center*
Rachel Tumin, *Ohio Colleges of Medicine Government Resource Center*
Bo Lu, *The Ohio State University*
Marcus E. Berzofsky, *RTI International*
Chengzhou Zhang, *Ohio Colleges of Medicine Government Resource Center*

Due to limited sample size, often health survey analysis results are confined to macro geographical units such as states and national regions because direct small area estimates are often too unreliable at the micro levels. However, with the implementation of the Affordable Care Act (ACA), there is a growing policy demand for determining health risk, health care coverage and health care utilization variation within and between small areas within states. In Ohio, the ACA policy research demand is to provide baseline health information at the county level (n = 88). To meet this demand, the authors applied small area estimation techniques (SAE) to the Ohio Medicaid Assessment Survey (OMAS). Various techniques were piloted with the authors finally choosing synthetic estimates to modeled estimates upon 8 key variables of policy interest (i.e., uninsurance status, insurance purchased from the health care exchange, self-rated health status, etc.). Attention was given to the selection of covariates for these SAE estimates across health policy variables. Consideration was also given to the dual-sample frame setup of the OMAS, which is a complex designed household survey with a n-size of 42,876, with 26,423 (61.6%) being cell phones. Not unlike national surveys, the OMAS n-size provides a robust sampling base and allows for estimates within State regions (i.e., Appalachia, metropolitan, rural non-Appalachian, and suburban areas) and for urban counties, but falls short of the minimum necessary respondents to make reliable direct survey estimates for the smaller counties. In this paper we discuss SAE challenges when analyzing policy data within small counties. To overcome these challenges the authors developed county level direct, synthetic, and a composite estimator as a weighted average of the modified direct estimator and synthetic estimator. The authors used the bootstrap method to estimate the MSE for the synthetic estimates.

State-level Estimates from the NHIS Restricted Data: Analysis to Support States Implementation and Evaluation of the ACA

Joanna M. Turner, *University of Minnesota*

The Patient Protection and Affordable Care Act (ACA) involves significant changes in health insurance coverage and health care systems across the United States. Many of the provisions of the ACA are being implemented by states and state specific factors such as health insurance coverage, access, use, and affordability will influence both implementation and impact of the law. As a result, state program administrators and policymakers need reliable state-level data to fully understand the impacts of the ACA over time. The National Health Interview Survey (NHIS), the nation's primary survey for monitoring the nation's health, provides a strong base for informing both the implementation and

evaluation of the ACA. The NHIS provides national estimates for a variety of health indicators, but only selected estimates are available for states with sufficient sample size. This study expands the role of the NHIS for state-level analysis of key health care outcomes. We submitted a proposal to the National Center for Health Statistics to work in a Census Bureau Research Data Center (RDC), for access to state identifiers, to develop state-level estimates for ACA relevant indicators and to make these publicly available. The NHIS state identifiers are merged with the University of Minnesota's Integrated Health Interview Series, a harmonized version of the NHIS. We explored creating a state-based weight and discuss challenges of working with the restricted data. We provide an overview of working in an RDC, as well as the disclosure process. We demonstrate how to access the 2011 to 2014 NHIS state-level estimates through an online interactive data tool and discuss gaps in the data that remain for assessing health reform. As part of this project, we plan to create an on-going series to support the tracking of health outcomes for states over time.

Measuring Survey Behavior of Smartphone Users
Samantha Luks, *YouGov*
Rebecca Phillips, *YouGov*

Most of American adults have a smartphone and an increasing number of them do not have any other form of high-speed Internet access at home. A successful adaptation of web surveys requires not only readjusting the format of the questions to a smaller display that allows for direct interaction (i.e., touchscreens) but also accounting to the fact that surveys on smartphones are more likely to be taken in contexts with more environmental stimuli, and therefore with greater chances of distractions and interruptions. In this paper, we study the attention span of survey respondents on smartphones by comparing their responses and behaviors to a matched sample of computer survey respondents. In a collection of experiments, we manipulate both the announced and the actual length of the interview, and measure the reliability of their responses using a large battery of checks. Additionally, we measure a number of survey metrics, including breakoffs, taking temporary breaks, rotating devices, and switching devices. The contribution of our paper is two fold. First, we provide practical guidelines for the design of web surveys that are possibly taken on a multiplicity of devices. Second, we make a methodological contribution by presenting it in an setting designed to recover the optimal responses.

Attitudes About Immigrants in a Shifting Cultural Landscape
Dan Cox, *Public Religion Research Institute*
Rachel Lienesch, *Public Religion Research Institute*

Over the last couple decades, the amount public opinion research conducted in the U.S. has exploded. Dozens of domestic polling firms and survey houses regularly add to an already large pool of national attitudinal and demographic data. However, despite the increasing availability of high quality national survey data, there is a dearth of public opinion research that is conducted among smaller geographic units, such as states and major metropolitan areas.In an effort to address this shortage, Public Religion Research Institute (PRRI) launched the American Values Atlas (AVA) in 2014. The AVA is a dynamic interactive online map of the details the cultural, religious, and political complexities of the American landscape. The AVA is based on more than 50,000 telephone interviews among a random sample of American adults conducted over the course of the year. Because of its large sample size and wide scope, the AVA provides users with a unique opportunity to explore the religious, political, and demographic diversity across all 50 states and in 30 major metropolitan areas. Additionally, the project provides a rare portrait of smaller religious communities and ethnic groups that are difficult to capture in smaller public opinion surveys.In this poster presentation we will present new findings from the 2015

AVA covering a wide range of topics, including religious and racial diversity, political identity, and views about important public policy questions such as immigration. We will pay particularly close attention to state-to-state variation in attitudes on these important questions and examine state-level differences in religious and racial diversity. The poster, like the AVA project overall, provides a unique window into the rising racial, ethnic and religious diversity in the U.S. and explores how the new cultural landscape impacts views about immigrants and immigration policy.

Coding Open-end Responses: Man vs. Machine
Besheer Mohamed, *Pew Research Center*
Elizabeth Sciupac, *Pew Research Center*
Dennis Quinn, *Pew Research Center*

As society progresses to the point where many tasks can be automated, the question arises: how does that affect survey research? Already there is a substantial amount of programming and automation that goes into call lists and interviewing, but the analysis is still done, primarily, by human researchers. One level of analysis, though – open-ended response coding – may actually lend itself to automation. In recent years, the field of machine learning – the study of how machines can learn to emulate human behavior – has produced a wide variety of algorithms that can analyze the content of some text documents with near-human accuracy. One of the main tasks for these algorithms is categorizing documents by topic – a task that survey researchers regularly undertake with open-ended survey responses. This paper takes up the question of how and in what circumstances these algorithms can be used in place of manual coding, what their weaknesses are, and how they can enhance the performance of human coders. To date, this capability has not been applied widely in the survey research community for several reasons. First, open-ended responses present a unique challenge to machine learning algorithms because they tend to be short, containing fewer words than most documents that these algorithms were built for. Second, open-ended questions often contain poor grammar and spelling, a challenge for computers. To assess the effectiveness of machine coding, Pew Research compared manually coded open-ended questions to those coded using a variety of machine learning techniques. We found that machine learning algorithms can be a valuable tool for increasing the efficiency of human coders.

Targeting Asian Subpopulations in a Cell Phone Random Digit Dial Survey: Using Census Data to Improve the Usefulness of Billing ZIP Codes
Michael Jacobsen, *RTI International*
Matt Jans, *University of California, Los Angeles*

Geographic targeting of racial and ethnic populations with cell phone samples is difficult due to the relative lack of geographic information on the sampling frames. The billing ZIP code of the cell phone account assigns cell phone numbers to a geographic area, and are a potential way to target subpopulations, thereby reducing coverage bias. However, about 50% of records with billing ZIP codes report living in their billing ZIP codes (Dutwin, 2014). This limits the ability of billing ZIP codes to target subpopulations. By combining billing ZIP codes to Census geographic regions, using American Community Survey (ACS) data may improve the utility of billing ZIP codes in targeting subpopulations. This study will examine the ability of Census data in augmenting the efficacy of billing ZIP codes in targeting a subgroup in a cell RDD survey. Specifically, this study will examine how appending ACS data can improve the ability of billing ZIP codes in targeting the Asian subpopulation in a cell phone survey. The USPS ZIP Crosswalk Files produced by the Department of Housing and Urban Development estimate

the proportion of residential addresses in a Census region that belong to each ZIP code that crosses that region. The proportion of Asian households in each ZIP code were calculated using these files and the 2013 5-Year ACS Asian household tract-level counts. The ZIP codes were then categorized into quartiles based on their proportions of Asian households. Cell phone respondents to the 2015-2016 California Health Interview Survey who both indicated they belonged to an Asian ethnicity and had a billing ZIP code were matched by billing ZIP code to the quartiles. Initial findings show that 75% of these Asian respondents were in the top 50% of ZIP codes. However, non-Asian respondents with billing ZIP codes were evenly distributed among the quartiles.

As Time Goes By: The Impact of Time Lapse Between Screening and Interview Recruitment of Older Adults in Survey Research

Sara A. Walsh, *NORC at the University of Chicago*
Lauren Sedlak, *NORC at the University of Chicago*
Daniel Lawrence, *NORC at the University of Chicago*
Anna F. Wiencrot, *NORC at the University of Chicago*

In both survey and gerontological research, gaining cooperation with adults 50 or older is a challenge. Previous research indicates that response rates decrease as age increases (Groves and Couper 1988; Herzog and Rogers 1988; Redpath and Elliot 1988, Murphy 2004). As Americans are living longer, the response rates of this population are of increasing importance. The National Social Life, Health, and Aging Project (NSHAP) is a longitudinal, population-based study of health and social factors in older adults. The original cohort (currently aged 67- 95) was recruited in 2005 (Wave 1) and a new cohort (aged 50-67) was recruited in 2015 (Wave 3). The Wave 1 cohort was identified through in person screening from February to November 2004 and then re-contacted over a year later, in July 2005, for recruitment into NSHAP (O'Muircheartaigh et al 2009, Smith et al 2009). The Wave 3 cohort was also identified by in-person screening, but there was no delay between screening and recruitment into NSHAP, a protocol we called "Screen and Go." By comparing data from Wave 1 and Wave 3, we intend to examine what impact differences in recruitment methods have on response rates in older adults. We will focus on survey completion rates, time to survey completion, and respondent characteristics to answer the question about whether time elapsed between screener and invitation to participate drives response rates in older adults. Little research has been done examining time between screening and recruitment for in-person interviewing, and our findings will help other studies looking to include older adults in their sample as well as lend insights into general survey recruitment methods, especially household screening.

Correcting Biases in Auxiliary Data to Produce Better Estimates

Masahiko Aida, *Civis Analytics*

This paper examines methods to correct for biases in auxiliary data (ex. voter list, consumer database) using higher quality reference data (ex. Census data) in order to produce better ratio estimates (regression estimates). Survey practitioners are increasingly eager to use auxiliary data (frame information) for statistical inference. This is particularly true in light of a rapidly deteriorating survey environment including budget cuts in government agencies and more stringent laws (ex. TCPA) driven by privacy concerns. Some innovative practitioners started using population frames such as voter lists and consumer databases for survey inferences, especially for political campaign work after 2012 (Nickerson & Rogers, 2014). The predictive scores that campaigns and some partisan polling firms use are in essence, regression estimators derived from both survey data and auxiliary data. Outside of

political campaigns, such model based approach to obtain more accurate (in terms of MSE) estimates are already in use in the fields of small area estimation methodology (ex. the US Census' SAHIE program). In theory, with a good frame data and a good model, researchers should be able to accurately estimate population statistics. However, frames such as voter lists and consumer databases contain systematic error of omission, inclusion and duplication. People with unstable living condition may not be captured. Deceased people may remain in the file after they are long gone. A young person may be double counted at her parent's residence and at her new address.All of these errors in the frame, if unattended will manifest as biases in the final estimates. The author will compare different correction methods to remove biases from auxiliary data and evaluate the impact on total error (MSE) using 2014 midterm election and insurance coverage status survey data.

"Like Therapy" but to No Effect: Repeated Interviewing Does Not Impact Attitudes or Experiences in the Case of Fertility Treatment Choice

Alexandra Cooper, *Duke University*
Amanda Flaim, *Social Science Research Institute*
Kathryn Flynn, *Center for Patient Care & Outcomes Research*

Collecting information from human subjects creates a problem for those seeking to understand attitudes and behaviors; the experience of being studied can affect outcomes of interest, and such effects can be difficult to assess. This poster examines a study of infertility treatment choices that deploys sustained and repeated line of inquiry with a subset of couples who are exploring treatment. Such a study could influence the views, interactions, and behaviors of participants, potentially altering its findings. Yet a division between participants with regard to the extent of their involvement in this study provides a unique opportunity to test the degree to which extended reflection and intensive interviewing may affect research findings. In this research, we compare two subsets of participants in the study in question to determine whether participating in intensive and repeated interviewing affected participants' reactions to their experiences as they consider fertility treatments. Comparing these two groups – one group that participated in interviews an average of once every two months, and one group that completed only two sets of self-administered surveys, twelve months apart – identifies few if any detectable differences in the experiences of subjects across the two arms of this study. This poster presents data comparing the two sets of study subjects and their experiences and attitudes, and discusses the reassuring finding that the process of comparatively intense data collection – an important tool in understanding human behaviors and attitudes – does not necessarily alter their experiences or attributes.

Using a Non-probability Sample to Understand the Ongoing Needs of Superstorm Sandy Victims

Timothy Gerard Tracey, *Monmouth University Polling Institute*
Patrick Murray, *Monmouth University Polling Institute*

The Sandy Recovery Panel Survey has been tracking the progress of approximately 500 New Jersey residents whose home was significantly damaged by Superstorm Sandy. Panel participants were recruited using a variety of non-probability methods, and interviewed both online and by phone annually over the past three years. Despite the lack of a probability sample to statistically extrapolate the sample to represent all Sandy victims in New Jersey, the data have provided a unique perspective on the recovery process. The results of this study have proven useful for advocates and policymakers because they fill a glaring data gap about victims' experiences. The use of non-probability sampling was

necessitated by three main factors: the lack of a publicly available list – or even reliable estimates – of residents who suffered significant damage; the disperse geographic impact of the storm; and the difficulty in locating displaced residents. The study team deployed a combination of purposive and snowball sampling, using FEMA damage assessment maps to identify the hardest hit neighborhoods in order to recruit participants in person, attending victim resource meetings and hearings, and working with local officials to publicize the survey. This poster will review the methodological challenges of recruiting the panel, as well as maintaining a representative sample of the original panel participants through each wave of the study and how the panel compares to the few known demographics of the larger population it represents. In addition, the utilization rates of government recovery assistance programs will be examined to assess whether the panel study responses can be used as a check on official government statistics. Finally, we will highlight results from the 3-year panel that illustrate relationships between home-displacement status over time and key factors such as ongoing social and financial needs and mental health issues.

Hello, Are You There? A Look at the Impact of Incentives in a Longitudinal Evaluation

James P. Cooper, *Decision Information Resources, Inc.*
Lenin G. Williams, *Decision Information Resources, Inc.*
Scott Peecksen, *Decision Information Resources, Inc.*
Sylvia R. Epps, *Decision Information Resources, Inc.*
Lee Robeson, *Survey Management, Inc.*
Jo Anna Hunter, *MDRC*

There is a strong literature on the importance of providing incentives for survey completion and in particular for longitudinal studies. With longitudinal studies, the first survey experience will set the tone for any follow-up efforts. Further, Jäckle and Lynn (2008) found that payments at multiple waves, significantly reduced attrition in all waves. For the proposed poster, data from three waves of a multi-site experimental evaluation will be examined to determine the impact of incentives on subsequent waves. Specifically, authors will examine patterns of responsiveness of registrants for programs targeting low-income and disadvantaged populations offering subsidized employment in several cities in the U.S. Data will include demographic variables, group status, site, incentive amount, number of attempts (CATI and field), and the number of days to complete at each wave. Across the waves, data are available for ~7,000 people. The specific research questions to be addressed are: 1. Did Wave 1 early responders (completed within 30 days) respond at Waves 2 and 3, and what was their average days to complete? 1.a. Did this vary by group status, site, or general? 2. What is the attrition rate between waves for Wave 1 responders? 2a. Did this rate vary by group status, site, or in general? Answers to these questions will determine whether the responsiveness in Wave 1 impacted responsiveness at subsequent waves. Data is drawn from three waves of surveys conducted in 4 cities under contract to MDRC for the Subsidized and Transitional Employment Demonstration (STED) program, funded by the U.S. Departments of Health & Human Services and Labor.

Bringing Fair Market Rent Surveys into the 21st Century – Evaluating the Effectiveness of MSG's Email Flag on an Address-based Sample Design

James Dayton, *ICF International*
Thomas Brassell, *ICF International*
Vanessa Cooper, *Housing Authority of the City of Alameda*
Ron Dion, *Housing Authority of the City of Alameda*
Richard Williams, *Vermont State Housing Authority*

Annually, the U.S. Department of Housing and Urban Development (HUD) uses the American Community Survey (ACS) to calculate the fair market rent (FMR) that establish the rent subsidy levels for people in need rather than actual market rent data. Public Housing Authorities (PHAs) may appeal these rates by administering a FMR survey to collect monthly rent and utility costs from a random sample of households using an address-based sample (ABS) design. FMRs are calculated using the data from two bedroom renters that moved within the past two years, resulting in very low eligibility rates (traditionally one to three percent). With a HUD-defined target of 200 eligible completes, required sample sizes often make these surveys cost-prohibitive for PHAs. A move to a fully web-based administration could substantially reduce costs and allow PHAs to pursue FMRs that are more reflective of current market rents. Drawing a random sample using Marketing Systems Group's (MSG) email flag (with an expected 15%-35% match rate depending upon location), we will administer the FMR survey via email with personalized web link, with a mail control group, to selected respondents in a HUD-defined metropolitan statistical area (MSA). The goal of our research is three-fold. First, we will investigate the effectiveness of MSG's email flag in providing a valid email address. Second, we will conduct a cost/benefit analysis assessing the cost differences and eligible completes received between the respective modes. Lastly, we will assess for mode effects. In 2016, approximately 289 MSAs experienced a decrease of $25 or more to the established rental subsidy – a substantial amount for individuals in need. Finding alternate, cost effective means of conducting FMR surveys would allow more PHAs to pursue HUD appeals, allowing for an opportunity to obtain FMRs that will better serve people in need of housing support.

Asking About Religion: Other, Atheist or Agnostic?

Devin Van't Hof, *University of Nebraska-Lincoln*
Kristen Olson, *University of Nebraska-Lincoln*
Jolene D. Smyth, *University of Nebraska-Lincoln*

Self-reported religious identity continues to change in the United States. Fewer respondents are identifying as Christians (Catholic or Mainline Protestant) and more respondents are identifying as either minority religions (Muslim and Hindu) or as 'unaffiliated'; with the largest increase among respondents who identify as 'unaffiliated' (Pew Research Center: Religion and Public Life, 2015). These 'unaffiliated' respondents are often grouped into one category instead of being separated out into categories such as atheist or agnostic. For example, when the General Social Survey (2010) asks about religion it presents the choices of Protestant, Catholic, Jewish, other religion, or no religion. By combining all 'no religion' respondents into one category researchers may be missing appreciable differences between respondents. This papers begins to examine the differences between responses and respondents when the additional religious options of 'none', 'atheist', and 'agnostic' are provided in a religious affiliation question compared to the sole option of 'none'. A nationally representative mail survey of 6000 sample members was conducted during spring 2015 (AAPOR RR1 = 16.7%). All sample members were asked "What is your religion, if any?" with a random half of the sample members given the shorter answer

stem of "None, Catholic, Jewish, Muslim, Protestant, and Other" and the other random half given the longer answer stem of "None, Atheist, and Agnostic, Catholic, Jewish, Muslim, Protestant, and Other". There are significantly more unaffiliated (answering either 'none', atheist', or 'agnostic') respondents when the atheist and agnostic options are included. Additionally, preliminary analyses indicate that older and politically moderate respondents differentially select an affiliated vs. unaffiliated category when presented with "atheist" and "agnostic" compared to not having those options presented. The paper will conclude with implications for asking about religious affiliation in surveys.

Respondent Selection in the Behavioral Risk Factors Surveillance System

Robert Tortora, *ICF International*
Naomi Freedner-Maguire, *ICF International*
Samantha Vincent, *ICF International*
Eric Miller, *ICF International*

Three adult selection procedures, random, next birthday and last birthday, are compared for the 2015 Rhode Island and 2016 Connecticut Behavioral Risk Factors Surveillance System surveys (BRFSS) landline portion of the sample. Even though the birthday rules are quasi-random selection methods they are worth evaluating since they may impact response rates positively as well as not introduce bias into the substantive data. Since the repeatability of some scientific results has recently come into question we choose to conduct the comparison in two different states. The data was collected in Rhode Island in May, June and July of 2015 and will be collected in Connecticut in January, February and March of 2016. Fifty percent of the sample uses the operational rule and twenty-five percent of the sample uses the next birthday and the remaining twenty-five percent uses the last birthday rules. The rules are compared on several dimensions including response and refusal rates, length of interview and demographics and some key health variables including Gender, Age, Education, Ever smoker, Current smoker, self-rated health, Hypertension, Asthma and Diabetes. Preliminary results from Rhode Island indicate no significant differences except for Age.

Modeling Response Propensity in a Longitudinal Study of Low-income, Young Persons

Julie Pacer, *Abt SRBI*
Kelly Daley, *Abt SRBI*

Enrollment into a longitudinal study includes collecting detailed respondent information and secondary contact information to ensure high response rates to a later follow-up survey. These contact data are vital to locating participants of a longitudinal study that are low-income and young, a population that may be mobile. In an experimental study with random assignment, additional data are collected at enrollment which may also be useful in locating study participants to administer a follow-up survey. This research will explore ways to adapt the enrollment and survey approach for similar hard-to-reach populations in experimental, longitudinal surveys. We will develop a response propensity model using sample frame data such as the quantity and quality of contact information provided as well as other factors like gender, age, region, and treatment group status to understand whether certain enrollment data are predictive of survey completion. Additional analyses will explore whether these data may predict the level of effort required to reach the study respondent.

The Effects of Compensation Levels on Sample Demographics on Amazon's Mechanical Turk

Nicholas Hatley, *Pew Research Center*
Kyley McGeeney, *Pew Research Center*

Nick Hatley and Kyley McGeeney, Pew Research Center Amazon's Mechanical Turk is an online labor market where "workers" are paid for completing tasks. "MTurk" has become an increasingly popular tool for academics and social scientists that wish to conduct experimental research. Mechanical Turk provides significant cost and time advantages to other sample sources that may be available. However, the validity of MTurk samples has been heavily questioned, and the demographic make-up of who completes tasks on MTurk is difficult to study. There have been recent efforts to better understand the demographic make-up of MTurk. Panos Iperirotis, a professor at the Department of Information, Operations, and Management Sciences at Leonard N. Stern School of Business of New York University, created a service that continually tracks demographic data on Mechanical Turk workers by running continuous surveys that award workers $.05 for answering five demographic questions. While an enormous contribution to better understanding the demographic characteristics of those willing to complete tasks on MTurk, these data may not be representative of what other researchers can expect to find. Preliminary research by the Pew Research Center is supportive of this hypothesis. A compensation level of $.05 is out of line with what most researchers normally offer to their research participants. This research will report the differential effects that a number of compensation levels may have on samples collected through Mechanical Turk. The presentation will also feature subgroup analyses that may be useful to researchers that wish to better understand how overrepresented subgroups may impact treatment effects.

National Estimates of Sexual Minority Women Alcohol Use through Web Based Respondent Driven Sampling

Deirdre Farrell Middleton, *ICF International*
Ronaldo Iachan, *ICF International*
Naomi Freedner-Maguire, *ICF International*
Karen Trocki, *Alcohol Research Group*
Christian Evans, *ICF International*

We describe a methodology that ICF developed for on-line respondent driven sampling (RDS) recruitment and telephone data collection in order to assess alcohol use in sexual minority women. Risk behaviors and protective factors in sexual minority women cannot be effectively studied in national surveys due to insufficient sample sizes, and many studies of this population are limited to small geographic regions. This methodology allows for national estimates of key risk behaviors and protective factors. The study includes novel methods for selecting seeds for the RDS sample from the National Alcohol Survey. Seeds and their eligible recruits are given recruitment coupons via email which they are incentivized to forward to contacts in their social network who may be eligible. Interviews are scheduled and conducted via phone. We will discuss the outcomes of the initial pilot testing this approach. The pilot includes 10 seeds and the subsequent recruitment chains which occur over the course of 2 weeks. If data are available, we will discuss the results of the full study which will include an anticipated 1,000 interviews. We will discuss both the effectiveness and challenges identified with the methodology, including efforts to prevent incentive collection scams, as well as present key estimates for risk behaviors and protective factors related to alcohol use in sexual minority women.

Augmenting Traditional Estimates with Non-designed Data: Modeled Estimates of Consumer Confidence

Robert Montgomery, *NORC at the University of Chicago*
Martin Barron, *NORC at the University of Chicago*

The era of Big Data has arrived for survey research and brings with it a variety of opportunities and challenges. One approach to taking advantage of Big Data is to integrate it with probability-based survey (i.e., traditional) data to create a blended estimate. A model estimate that blends traditional and non-designed data sources may have a number of advantages over estimates based on either source alone. Blended estimates may be timelier, may allow for estimation to smaller geographic areas, and may be more cost effective than estimates based on traditional data alone. At the same time, leveraging the strength of probability-based survey estimates in terms of known variance and bias properties can enhance and improve estimates based on non-designed data alone. Preliminary research conducted by the authors illustrated the potential of this approach in modeling unemployment (combining data drawn from survey estimates, job board listings, social media data, and search engine query aggregations). In this poster we expand upon our earlier research by modeling consumer confidence, combining survey-based estimates with non-designed data. Our research explores the overall comparability of these modeled estimates to the survey-based estimates, their ability to allow appropriate small area estimates, and their general cost-effectiveness and timeliness.

Social Norms, Identity-relevant Media and Pushback: Attitudinal Responses to Controversial In-group Claims About Health

Aaron S. Veenstra, *Southern Illinois University*
Zachary A. Sapienza, *Southern Illinois University Carbondale*
Lindani Mbunyuza-Memani, *Southern Illlinois University Carbondale*
Benjamin A. Lyons, *Southern Illinois University Carbondale*
Jin Lee, *Southern Illinois University Carbondale*
Stephanie Cheeyoun Kang, *Southern Illinois University Carbondale*
Steven Giannino, *Southern Illinois University Carbondale*
Alev Degim, *Southern Illinois University Carbondale*
Angela Anima-Korang, *Southern Illinois University Carbondale*

Recent studies have examined how social identity and identity-relevant media impact beliefs about health controversies, including the claim that vaccines cause autism. Specifically, religious, parenting, and environmental identities have been found to be especially relevant. Identity-related messages and media (e.g., religious broadcasting, parenting magazines) can enforce group norms, suggesting that anti-vaccine messages that contain an identity frame and appear in such media might be particularly influential on vaccine attitudes. When a particular social identity is primed, that identity's reference groups become highly accessible for cognition; thus, their norms become part of the attitude formation process. We hypothesize that exposure to an anti-vaccine message framed with a relevant identity will promote anti-vaccine attitudes and discourage trust and efficacy regarding traditional health care, an effect that will be heightened when the message comes from an identity-relevant media source. Data come from an experiment using a sample of 400 MTurk users. Participants were presented with an editorial opposing childhood vaccinations on the grounds that they are dangerous, and promoting parental choice. Norms presented in the editorial were manipulated to connect to Christian, parent, or environmental identities. For analysis, participants were scored as high or low identity salience for the condition they were in. Another factor manipulated whether the editorial was presented

in an identity-relevant online magazine (e.g., Parenting) or in Newsweek. Multiple pro-health attitudes were actually bolstered in high-identity salience conditions, and especially the identity-media condition, contradicting our hypotheses. Disaggregating the three identities, the effect was driven by parenting identity, while environmental identity results were similar but inconsistent. In contrast, religious identity contributed to some negative vaccine attitude outcomes. Results were concentrated among likely mediating variables for effects on vaccine attitudes, such as trust in doctors and health efficacy. Generally, higher identity salience resulted in pushback, suggesting contestation of normative expectations among high-identifiers.

Attention, Recall and News Coverage of Polling
Kirby Goiidel, *Texas A&M University*
Joe Ura, *Texas A&M Uniiversity*

In this study, we present experimental results investigating the effect of poll-centered news coverage on attentiveness to news stories, willingness to seek additional information, and recall of issue-based information. Specifically, we examine how the emphasis of poll results in the headline influences decisions to read a news story, the amount of time spent reading the story, willingness to look up additional information online, and recall of issue-based information. We expect to find that while poll-centered news attracts attention, it constrains attention to those candidates leading in the polls and it has little effect on recall of issue-based information. In the conclusions, we discuss the implications for democratic governance.

A General Survey Measure of Individual Listening Styles: Short Form of the Listening Styles Profile-revised (LSP-R8)
Eike Mark Rinke, *University of Mannheim*

An 8-item short form of the Listening Styles Profile-Revised (LSP-R8) for use in general population surveys is proposed. The use of survey measures of citizens' political talk and conversation networks has been a staple of political communication research for decades (e.g., Lazarsfeld, Berelson, & Gaudet, 1944) and some of the most productive recent debates in political science have been sparked by findings based on survey measures of political talk (e.g., Mutz, 2006). Their use continues to inform various productive line of research. In this methodological brief, I present a new scale for use in general surveys intended to allow for the measurement of an as yet understudied aspect of political talk: citizens' general listening dispositions. The best-validated measure of these dispositions, the Listening Styles Profile-Revised (LSP-R), measures an individual's general disposition to: (a) listen to understand emotions and connect with others (relational listening); (b) withhold judgment and consider all sides of an issue before responding (analytical listening); (c) focus on inconsistencies and errors during conversations (critical listening); and (d) see listening as a transaction aimed at solving concrete problems (task-oriented listening). However, the original 16-item self-report measure (LSP-16, Watson et al., 1995) and the revised 24-item scale (LSP-R, Bodie, Worthington, & Gearhart, 2013) is obviously too lengthy for use in general population surveys. The methodological brief reports the first successful validation of an adapted short version of the LSP. The proposed LSP-R8 is based on LSP-R items and was validated in a two-wave panel study of German undergraduate students (Nt1 = 176; Nt2 = 165; N t1&2 = 115). Tests of internal consistency reliability, test-retest reliability (four-week interval), and confirmatory factor analyses of the theoretical four-factor measurement model produced strong validity evidence for the LSP-R8, which is about 60% more economic in terms of completion time than the LSP-R.

How Can We Develop a Survey that Everyone Can Answer? Developing and Cognitively Testing a Survey for Individuals with Physical and Mental Challenges
Elizabeth Frentzel, *American Institutes for Research*
Ushma Patel, *American Institutes for Research*
Graciela Castillo, *American Institutes for Research*
Beth Jackson, *Truven Health Analytics*

The goal of this study is to develop an experience of care survey for Medicaid home and community based services (HCBS) participants in order to capture feedback on their experiences with program services and supports. This survey is intended to be a cross-disability instrument that assesses services for adults in several populations: individuals with physical disabilities; intellectual disabilities; cognitive impairments; developmental disabilities; severe and persistent mental illness; and older adults (>65 years) with limitations in activities of daily living and/or instrumental activities of daily living. There are currently many surveys that examine experiences of care, but typically these focus on a specific population. Other experience of care surveys simply do not include these population(s) or they use proxies to answer on behalf of the individuals. We conducted formative interviews across multiple disabilities to identify important domains, developed survey items, and cognitively tested the items. A challenge in the development process was to align the survey items with the Consumer Assessments of Health Providers and Systems® CAHPS standard response options (0 to 10, never/sometimes/usually/always), and standardized item format. However, cognitive testing results showed that some participants found the items cognitively burdensome, although many could understand and respond. For example, ratings that used the zero to ten scale were hard to understand and use for those with low or no numeracy skills. Similarly, we tested multiple styles of response options and found that "mostly-yes" and "mostly no" response options were easier to for individuals with a cognitive impairment to understand, thus enabling more individuals to answer the survey. Finally, the standard use of a time reference was identified as cognitively challenging for some of the participants. This presentation will review results from the cognitive testing, including the experiment, to identify and describe issues related to developing surveys for these populations.

Study of Statistics and Stakeholders
Barbara C. O'Hare, *Retired, U.S. Census Bureau*

Our complex information-driven society is dependent on social and economic statistics. Many who have a stake in these statistics are not statisticians themselves, but as consumers of data produced by statistical systems, deserve to have their voices heard. Whether administrators in health and human services, market analysts in the corporate world, survey data collection operations staff, or the survey respondents themselves, a range of stakeholder group are affected by decisions to collect and disseminate survey data. Transparency in public decisions has increased in recent decades. Stakeholder engagement theories and practices first emerged in the corporate for-profit sector. The principles do not directly apply to the not-for-profit and government sectors and there has been relatively little systematic study of stakeholder engagement that might apply to survey statistics. This presentation is based on a paper commissioned by the National Academies of Sciences, Engineering, and Medicine Standing Committee on Integrating New Behavioral Health Measures. Opinions and statements in the paper are solely those of the author and not necessarily adopted or endorsed by the National Academies of Sciences. I will present key concepts of stakeholder engagement and examples of current activity within the Federal statistical system, toward the goal of encouraging further research on engagement processes, their cost-benefit, and development of best practices within survey

organizations. Benefits of statistical system decisions that include stakeholders can include maximizing the utility of the data, lead to more efficient use of resources, and create an advocacy community to support the survey. I will discuss the case for stakeholder participation, stakeholder concepts and theories relevant in the non-corporate environment, tools for engaging stakeholders and some techniques currently used in Federal surveys, concluding with proposed next steps to understand how processes that include stakeholders operate on decisions and to build a resource on best practices in stakeholder engagement.

Unpacking the Meaning of 50%: Uncertainty and Predictive Ability in Subjective Life Expectancy Questions

Colleen A. McClain, *University of Michigan*
Sunghee Lee, *University of Michigan*
Qianyin Huang, *University of Michigan*

Subjective probability questions are increasing in popularity, addressing topics as diverse as health and mortality, elections, and economic outcomes. From a respondent's perspective, however, such questions may require significant cognitive effort to think both prospectively and probabilistically—comprehending the relevance of a future event and making a judgment that maps a vague impression onto a numeric probability. Prior work shows that while these questions can at times be strong predictors of future events, responses tend to be "heaped" at rounded numbers, losing precision. In particular, the report of a 50% probability presents vague information; does the respondent mean that the event is as equally likely to happen as not, or that he or she just doesn't know the answer (as a form of nonresponse)? These manifestations of epistemic uncertainty raise concerns from questionnaire design and analytic perspectives. Are certain types of respondents are more likely than others to answer 50%, and to do so for different reasons? Using data from the Health and Retirement Study, we investigate conditions under which respondents report "50%" when asked about subjective life expectancy; that is, the probability that they will live to be a certain age. The 2008 HRS followed up with those who answered 50% to determine whether respondents were uncertain about the chances or believed that there were equal chances, providing insight into the response process. First, we use this information together with demographic and psychosocial predictors to understand who is most likely to report a 50% probability, as well whether the response process varies across subgroups. Then, we link to data from the 2014 HRS to assess how such responses relate to one's mortality status six years later, as a direct assessment of the impact of such answers.

Complementary Facility and Population Surveys for Identifying Barriers to Health Care in Developing Countries

Aubrey Levine, *University of Washington*
David Phillips, *University of Washington*
Stephen Lim, *University of Washington*
Moses Kamya, *Infectious Diseases Research Collaboration*
Adoke Yeka, *Infectious Diseases Research Collaboration*
Gloria Ikilezi, *University of Washington*
Joseph Dieleman, *University of Washington*
Jessica Shearer, *PATH*
James Okello, *Infectious Diseases Research Collaboration*
Alexandra Wollum, *University of Washington*
Katya Shackelford, *University of Washington*
Emmanuela Gakidou, *University of Washington*

The Institute for Health Metrics and Evaluation (IHME) has conducted CAPI surveys in 20 countries assessing the operations and performance of health facilities. A multi-pronged approach to data collection is used. IHME's health facility survey modules target a variety of records and sources within establishments to collect data on infrastructure, personnel, supplies and equipment, finances, and services performed. Additional surveys are also fielded concurrently, including alternate data sources within the facilities' patient and catchment populations. Patient exit surveys are administered to assess patient perspectives and demand-side factors. Medical chart extractions and blood draws are also performed at facilities in order to collect data on patient health outcomes. Additionally, we take the unique approach of conducting complementary household surveys such that barriers to care can be identified through comprehensive models including both supply and demand-side factors. The process of data linkage and benefits of conducting complementary data collection alongside establishment surveys is presented utilizing an example of a constraints analysis in Uganda.

Embedding Archiving and Metadata into Data Collection: Policies, Procedure and Practices

Steven D. McEachern, *Australian National University*
Dina Neiger, *The Social Research Centre*
Janet McDougall, *Australian National University*

There is increasing demand in both academic funding agencies and government authorities for the provision of data from survey research projects in accessible formats in public repositories. The Office of Science and Technology Policy (OSTP) directed all agencies with research and development expenditures of more than $100 million that "federally funded scientific research are made available to and useful for the public, industry, and the scientific community ... [including] digital data". Similar policies have been established in the UK, Europe and Australia, providing a growing challenge for government agencies, universities and research institutions that are funded through public funding. This shift in expectations has signficant implications for the commissioners of survey projects - and for the field agencies and software providers who support the survey data collection process. The provision of data for public use entails a corollary requirement for adequate documentation about both the research methodology and the data itself. In order to meet these new mandates, commissioning agencies are therefore likely to have increasing demands for both documentation of data collection activities and final outputs, including the data files themselves and associated reports and metadata. This paper

therefore describes one effort to meet these requirements. The Australian Data Archive and the Social Research Centre in Australia have been collaborating on a project to develop procedures and practices for embedding data documentation requirements within the data collection workflow process, to enable automated capture of relevant documentation as part of the standard activities of the collection agency. The project explores the joint efforts of ADA and SRC to facilitate capture of documentation requirements for data deposit with ADA (as the research data repository), and provision of suitable metadata to support agency requirements for both access and data discovery (including the Australian National Data Service, DataCite, and data.gov.au).

Early Outreach Strategies with a H2R Population: Part I

David P. Getman, *Decision Information Resources, Inc.*
Natalia Ibanez, *Decision Information Resources, Inc.*
Sylvia R. Epps, *Decision Information Resources, Inc.*
Scott Peecksen, *Decision Information Resources, Inc.*
Ron Bass, *Approximetrix*
Jo Anna Hunter, *MDRC*

Longitudinal studies targeting hard-to-reach (H2R) samples face a twofold challenge: diminishing sample identification over time and outdated contact information. To combat this, early outreach efforts, study branding, and regular contacts are often employed. Still, important questions remain about whether early outreach is effective in maintaining a current database of sample-member contact information for survey launch or in sustaining sample identification within a longitudinal study.   The proposed poster will focus on an expanded tracking effort that sought to achieve two goals: (1) obtain updated contact information, and (2) ask four questions about material hardship. Respondents who completed received an incentive for completing. At 12-months and 18-months post random assignment, sample members received a letter with $2, encouraging them to update their contact information. We will present the various tracking efforts in detail and then using data about the responsiveness to these efforts, we will address the following: What is the responsiveness to the different tracking efforts?Did responsiveness vary by group status? Data are from a 6-month early outreach survey conducted under contract to MDRC for Grameen America, a microcredit/microfinance group-lending program operating in the New Jersey area. This study is a replication of an international effort. Given this sample includes a very specific population, we believe the results will be very useful in informing other large-scale evaluations on an understudied topic and population in the U.S. Additional analyses are planned to determine whether these early efforts impacted responsiveness at the upcoming 18- and 36-month follow-ups.

Exploring the Feasibility of Using Facebook for Surveying Special Interest Populations

Chanyoung Lee, *University of South Florida*
Si-Won Jang, *University of South Florida*

It is common that a survey aims to collect responses from specific population groups such as people with certain interests, people living in a certain area, people affected by a certain event, etc. In general, this kind of survey uses a registry or commercially available database to reach the targeted population in an effective and efficient manner. However, it is a challenging task as the registry or database information is often limited and expensive. In addition, primary available contact information is a phone number, which also limits the mode of survey to phone calls. During the last decade, web

survey has become much more common as telephone surveys have experienced many challenges. In this study, a web-based survey with a multi-method sampling approach was adopted to conduct a motorcyclist survey in Florida. Motorcycles represent less than 3 percent of registered motor vehicle s in Florida. The survey link was distributed through the following methods: Address-Based Sampling (ABS): Postcards with an invitation to complete the web-based survey were sent to a random sample of endorsed motorcyclists. In total, 30,000 addresses were sampled from the Florida DHSMV database of endorsed motorcyclists. Each sampled address was sent two invitation postcards over a period of two weeks. Facebook: An advertisement that targeted motorcycle riders ages 18 and older was placed on Facebook for 10 days to advertise the web survey. To increase participation, incentives were offered through drawings. Participants could enter to win a motorcycle safety-related prize, including a motorcycle jacket or a motorcycle helmet. In total, 3,817 survey responses (including partially-completed surveys) were collected. The study explored the cost-effectiveness of Facebook advertisement compared to ABS in terms of reaching the targeted population and discussing the potential coverage errors by age.

Using Modeled Auxiliary Variables to Reduce Bias and Variance in Political Polling

David Martin, *Civis Analytics*
Michael Sadowsky, *Civis Analytics*

Response rates in electoral polling continue to decline, threatening an increase in non-response bias. At the same time, many means to mitigate this bias are increasingly expensive. In short, obtaining accurate estimates of public opinion and forecasting elections is only getting harder, as some failures in polling in 2015 (namely in the UK and Kentucky) have shown. One potential bright spot in this picture comes from the rise of big data and machine learning techniques that can bolster public opinion measurement. To the extent that auxiliary variables exist that are known for the full population and are associated with variables of interest we can reduce bias and variance by adjusting samples to match that population (Little, 2005). We show how this technique can be used in the US political context using voter registration rolls. To illustrate this technique, we produce simulated samples and weighted estimates based on surveys conducted in advance of US Senate elections in 2014. We produce stratified bootstrap samples under two conditions: In the first condition we use modeled response propensity to simulate a typical RDD registered voter sample after nonresponse. In a second condition, we allocate sample to represent an electorate defined by age, sex, race, partisan support and vote history data that is either drawn directly from voter registration rolls or modeled from them. This simulates a sample representative of the modeled electorate sampled inversely proportional to response propensity. We weight each sample either to be representative of ACS demographic estimates or to be representative of the voter file on modeled demographic and behavioral indicators. Through this we show decreases in variance and bias compared to a simulated RDD sample weighted to ACS.

Economic Productivity and the Legitimation of Income Inequality: Evidence from a National Sample of the U.S.

Jonathan Kelley, *University of Nevada*
M.D.R. Evans, *University of Nevada*

Comprehensive data on public beliefs about the legitimacy of income inequality gathered from large, representative national sample surveys in many nations show: (1) broad agreement on the legitimate pay of low-status, ordinary jobs, (2) agreement that high-status, elite occupations should be

paid more than the minimum, but (3) disagreement over how much more they should get. This disagreement is linked to politics and social structure, with older, high SES, politically conservative respondents preferring markedly higher pay for elite occupations, but not usually preferring lower pay for ordinary jobs. In all this the US is, broadly speaking, typical. We will extend this flourishing line of research with new data from a national sample of the US (N > 1500) incorporating a new, multiple item measure of productivity. Judging from preliminary results from our national sample of Australians (N>1500) and a sample of US students (N> 180) , productivity can be measured reliably and has a very strong effect on the pay thought legitimate for elite occupations, far stronger than political preference or social structure. A doubling of productivity justifies something like a 50% increase in pay. This implies that productivity differences are a major factor in the legitimation of inequality in general. It also implies that ordinary levels of productivity growth in western nations (typically around 2% or 3% a year, more in some economic sectors and less in others) lead in time to changes in the pay structure ordinary people find legitimate. Insofar as productivity growth is greater in occupations requiring advanced skills and university based training, this implies an increase over time in the amount of occupation based income inequality that the public finds legitimate. But where productivity growth is slower (for example, in administration), it implies a more egalitarian distribution of income is legitimate.

Designing Probability Samples to Understand Treatment Effect Heterogeneity in Survey Experiments

Elizabeth Tipton, *Columbia University*
David Yeager, *University of Texas at Austin*
Ronaldo Iachan, *ICF International*
Barbara Schneider, *Michigan State University*

For those familiar with survey-experiments, the answer to questions regarding the ability to generalize findings is straightforward: use a probability sample. Statistical theory is clear here – this dual randomization procedure allows for the unbiased estimation of the average treatment effect in a well-defined inference population. As we will argue in this paper, however, while answering if these interventions result in differences on average, this design does not necessarily help researchers answer equally pressing questions regarding regarding how variable the impact of the intervention is, as well as if the impact varies across important subgroups. These questions align with three separate parameters to be estimated – the average, the variation in effects, and differences in average effects between pre-specified subgroups. As we will argue, combined these three parameters allow for the development of a theory of the causal mechanism. This presentation offers a new approach to the development of probability samples in survey-experiments, resulting in a study design that allows for all three estimands to be estimated precisely. To do so, the method requires researchers to begin not only by defining an appropriate inference population but also potential treatment effect moderators. In many cases, measures of these moderators will need to be developed, and then once developed, these measures are used to stratify the population. Finally – and of particular focus in small to moderate sized experiments – the final stratification design has to take into account concerns with statistical power for each of these estimands. In order to account for these competing concerns, we propose an approach for determining the optimal stratum allocation. Throughout, we situate this discussion in an example based upon the development of the probability sampling plan for the National Growth Mindset Intervention study, which includes nearly 80 schools througout the United States.

The Influence of Affect and Need for Orientation on Systematic and Heuristic Information Processing of Scientific Issues

Sara K. Yeo, *University of Utah*
Ye Sun, *University of Utah*
Jessica Houf, *University of Utah*
Meaghan McKasy, *University of Utah*
Erika Shugart, *American Society for Microbiology*
Emily Dilger, *American Society for Microbiology*
Joanna Urban, *American Society for Microbiology*
Sam Mandl, *American Society for Microbiology*

Information processing about scientific issues has significant implications for public opinion. Systematic processors analyze messages carefully, expending more cognitive resources, while heuristic processors use mental shortcuts to form relatively quick judgments. Here, we present a pilot study using the context of microbiomes and an experiment embedded in an online survey to test whether individual and message characteristics predict systematic or heuristic processing. The study of microbiomes is an emerging area of interdisciplinary scientific research that has significant environmental and human health implications. Examining information processing mechanisms about such issues is integral to understanding public opinion formation and dynamics. Undergraduate students enrolled in communication courses at a western university (N = 123) participated in the study for extra credit. The study employed a 2 (focus) x 2 (affect) x 2 (vocabulary) between-subjects design. After reading the short manipulation text (approximately 200 words), respondents completed questions, validated by Smerecnik, Mesters, Candel, De Vries, and De Vries (2012), intended to measure self-reported systematic and heuristic modes of information processing. The survey also included items tapping three dimensions of need for orientation related to microbiomes; concern, relevance, and interest, which were asked prior to exposure to the manipulation. Demographic characteristics (age, gender, education, and income) were collected for all participants. We focus our analysis on one aspect of the experiment, affect, and control for the other two. The affect manipulation exposed participants either to descriptions of the uses of microbiomes designed to elicit disgust or one designed to elicit a neutral affective response, i.e., no disgust. Results demonstrate that affect interacts with one dimension of need for orientation (interest) to differentially impact heuristic, but not systematic, processing. The implications of these preliminary results for information processing and public opinion of science are discussed.

The Mobile Landscape in the United States: How Much and How are Americans Using Their Smartphones?

Abhinav Agrawal, *SurveyMonkey*
Kelvin Law, *SurveyMonkey*
Noble Kuriakose, *SurveyMonkey*

Smartphone penetration in the United States has been steadily on the rise. According to the Pew Research Center, more than two in three adults now have a smartphone. However, researchers only have a limited understanding of how Americans use their mobile devices because of the challenges of collecting reliable data directly from smartphones. In this paper, we analyze many Terabytes of data to construct a large scale panel to glean mobile behavior and app usage, SurveyMonkey Intelligence. Our findings will address questions about how much Americans use their mobile devices and what Americans use their phones to do. We examine patterns of use such as mean time spent on phones,

usage of particular apps (including political and news apps) and their seasonal variations. In addition to topline figures, we examine differences across gender, age, race, and education.

Metadata Makes Better Data: Enabling Survey Design, Data Discovery and Re-use
Louise Corti, *UK Data Archive, University of Essex*

To some in the data business, modelling metadata is an obsession; to survey designers 'metadata' is an intrinsic part of everyday design, yet, may not be recognised using this terminology. Questions and variable both benefit from rich and standardised metadata if they are to be fully utilised: it is good for question design and data discovery, for comparison and harmonisation, and for creating future-proofed data documentation. It also makes for a more efficient recyclable data environment, where survey owners and producers can reduce the current manual efforts put into getting surveys in and out of the field. Finally rich metadata are the key to scaling up for exploiting 'bigger' data, enabling more robust discovery, visualisation and linking. The survey production landscape is still well behind in terms of exploiting metadata richness and recylability. In this talk I will explore some of the great work being done to help incentivise those working in the domains of the Social and Behavioural Sciences to embrace rich survey metadata. I will provide examples from the UK that focus on: persuading survey producers and survey software vendors to appreciate and smooth the metadata journey; the work being done to help researchers at the coalface understand why good metadata matters; work to retrospectively fit a range of complex study documentation into structured metadata for a common search portal; and outreach work with data repository managers to show why standardised metadata must be used for global data discovery and access. The Data Documentation Initiative (DDI) is the key, a useful schema specially designed for describing survey design and data. The trick is to appreciate what standard survey metadata can offer, show what is needed to 'get there', and stay away from unhelpful jargon, using community narratives to gain compliance with sharing re-usable information.

The Effect of Registration Mode on Customer Satisfaction with FEMA Registration Process
Kristin L. Brooks, *Federal Emergency Management Agency*
Brandi Lea, *Federal Emergency Management Agency*
Jessica Guillory, *Federal Emergency Management Agency*

The Federal Emergency Management Agency (FEMA) provides assistance to disaster survivors. Once a disaster is declared, the first step a disaster survivor must take in order to qualify for assistance is to register with FEMA. The registration process requires that applicants complete a 15-20 minute survey detailing their personal information and disaster related damages. To ensure the registration process is accessible to everyone affected by a disaster, FEMA accepts registrations via phone, internet, and in-person. In order to monitor customer satisfaction with the registration process, FEMA surveys a sample of applicants weekly throughout the application period. Currently, all customer satisfaction surveys are administered by FEMA surveyors via computer assisted telephone interviewing (CATI). The current project aims to explore whether satisfaction with the registration process differs across registration modes. Exploring satisfaction across registration modes may provide insight into consumer preferences for survey administration, and help identify whether the internet might be useful for administering FEMA's customer satisfaction surveys. Additionally, demographic variables such as age, income, and region of the United States will be included in the analysis to examine whether certain populations are more likely to use certain registration methods.

Exploring the Use of Classification Trees in Categorizing Survey Respondents on Perceived Survey Burden

Arcenis J. Rojas, *U.S. Bureau of Labor Statistics*
Scott Fricker, *U.S. Bureau of Labor Statistics*
Lucilla Tan, *U.S. Bureau of Labor Statistics*

Past research on interviewer observations about contacted sample units' attitudes and reactions ("doorstep concerns") to the survey request have found them to be predictive of final survey cooperation, and also respondents' perception of survey burden in the Consumer Expenditure Interview Survey, a national household panel survey. Additional burden items have been fielded but not yet analyzed since the previous studies. In this poster, we present the use of GUIDE software to perform classification trees analyses on 15 burden items to categorize respondents by their overall perception of survey burden. In addition, as an exploratory analyses to provide insights to a larger study examining the association of doorstep concerns observed in the first wave with final wave survey outcomes of interest, we examine how doorstep concern items associate with reported and predicted perceived survey burden.

SENSUS, Sensitivity and Safety: Randomization as a Best Practice for Understanding Respondents Ranking of Safety Concerns

Jasmine Wise, *Baylor University*
Emily Hunt, *Baylor University*

Researchers measured the attitudes and perceptions of Waco-McLennan County residents towards crime and victimization, safety and security, and community relations with the local police department. Resident assessments of police performance and evaluations of opportunities to improve police-community relations are included. Each respondent was asked several questions pertaining to attitudes toward various crimes and community problems, their feelings of safety, their assessment of police performance, and their evaluations of opportunities to improve police-community relations, as well as a variety of demographic questions. From February through April of 2015, a total of 1304 adult residents of McLennan County were interviewed using a Computer Assisted Telephone Interviewing (CATI) system. Using data collected from a longitudinal study of crime, victimization, and public safety in Mclennan County, the Baylor Center for Community Research and Development implemented a randomization approach when asking residents about serious problems in their neighborhood. Using the CATI system researchers randomized "most serious problem" questions on a crime and safety community based survey, we found differences in how respondents ranked pressing safety concerns from one wave to the next based on the randomization effect. This poster discusses how, using the randomization feature, survey researchers can reduce survey bias as well as respondent confusion.

Can Small Business Owners Describe the Health Insurance They Offer Their Employees Under the Affordable Care Act (ACA)?

David Kashihara, *Agency for Healthcare Research and Quality*

Effective January 1, 2014, the Patient Protection and Affordable Care Act (ACA) stated that all firms with 50 or fewer full-time equivalent (FTE) employees could purchase health insurance for their employees from health insurance marketplaces known as the Small Business Health Options Program (SHOP). Several questions were added to the Medical Expenditure Panel Survey – Insurance Component (MEPS-IC) to measure the impact of the ACA on the health insurance offered to employees. The MEPS-IC

is an annual survey sponsored by the Agency for Healthcare Research and Quality and conducted by the U.S. Census Bureau to collect information on the types and characteristics of health insurance offered by business establishments and governments throughout the United States. This paper describes the new ACA-related questions on the 2014 MEPS-IC and reports response patterns found in the data. For various reasons, response patterns indicated that some of the new questions were problematic for the respondents, showing that there was a pervasive lack of understanding or knowledge of some key terms that describe health insurance plans (e.g., "metal level" and "reference plan"). Two field studies were conducted on SHOP participants to clarify some of these ACA-related question response issues. The first field study contacted a sample of 2014 MEPS-IC respondents to gauge their knowledge of ACA-related terms as well as to clarify some of their questionnaire responses. The second field study aimed to determine why some SHOP participants, located in Federally Facilitated Exchange (FFE) states, reported offering multiple health insurance plans to their employees when only one plan was available from the FFE. The purpose, design, and results of these field studies are presented as part of this paper.

Comparing Representativeness in Online and Live Interviews Phone Surveys

Kevin Collins, *Analyst Institute*
Josh Rosmarin, *Analyst Institute*

With phone response rates dropping precipitously, survey researchers are increasingly turning to online interviews to measure attitudes. But how do samples of online and phone respondents differ, and which is more representative of the electorate? And when weighting to adjust for the non-representativeness of these samples, what are the implications for effective sample size? To answer these questions, we are conducting an experiment in two states where we randomly vary the means by which we attempt to interview registered voters, either attempting to conduct a live-interviewer phone survey (to landlines and cell phones) or sending voters mail to encourage them to participate in an online survey. We also randomly vary the incentive offered to mail recipients, promising either $2, $5, or $10 upon completing the online survey. Across these mode and incentive conditions, we compare demographics, political participation, and modeled political attitudes of survey completers, as well as the design effect that results from weighting each sample back to initial voter file sample based on these factors. We also compare other quantities of interest across incentive conditions including approval of elected officials, stated partisan preferences, self-reported media use, and Big Five personality traits.

Effect of Clarifying Instructions on Response to Numerical Open-ended Questions in Self-administered Surveys

Anil Kumar Chaudhary, *University of Florida*
Glenn D. Israel, *University of Florida*

In self-administered surveys, respondents must navigate through the questionnaire and interpret what each question is asking rather than assisted by an interviewer in a face-to-face or telephone interview. For an individual question, sound visual design helps respondents proceed through the response process. Visual designs are well tested by various researchers for their influence on reducing the measurement error and item non-response. In addition, Christian et al. (2007) and Christian (2007) found that adding a verbal instruction to the question's visual design significantly improved item response rates. Research on effect of clarifying instructions on visual design is self-administered surveys is, however, limited. In order to get further understanding of the use of clarifying information in self-administered survey questions, we designed an experiment which examined the effect of clarifying instructions (with/without) on responses to two numerical open-ended questions. Data for the study was collected using two questions from the 2015 client survey of Florida Cooperative

Extension Service (FCES). A web/mail mixed-mode design yielded 1,618 responses (51.4% RR2). The two questions used in the experiment were asked information about number of times using FCES this year, and years using Extension services, and clarifying instructions for two questions were "include 1 for the time that this survey asked about" and "write '1' if only this year" respectively. We found consistent evidence the added instruction clarifying how to respond reduced the percentage of missing and incorrectly formatted responses by 4.6 percentage points for the number of times clients used Extension and 7.5 percentage points for the number of years. Response mode did not affect these results.

Survey Experiment: Identity Ranking or Rating?
Jay Byron, *University of Massachusetts Boston*

Identity prominence (i.e., importance) is often measured in student populations by social psychologist using ranking scales. This hierarchical ranking of prominent identities may be a forced choice imposed by researchers. Based on data from a non-probability adult sample, this poster presents the findings from a web survey experiment evaluating the ranking and rating of prominent identities that are obligatory (e.g., mother, wife, worker). In this study an experimental group will be assigned rating scale questions whereas a control group will be assigned ranking scale questions. The purpose of the experiment is to determine if there is variability or equivalence of importance between obligatory adult role-identities. Compared to ranking questions, when respondents are asked to rate obligatory identities, do identities have equivalent importance or does a hierarchy of prominence emerge? This research addresses an important gap in survey research in identifying if rankings and ratings are equivalent measures of identity. Past research has demonstrated that ratings and rankings are not interchangeable measures of parental perception of important values possessed by children (Alwin & Krosnick, 1985). Additionally, Marks and MacDermid (1996) have argued, but not tested, that hierarchical ranking of identity prominence is a forced choice (see: Serpe, 1991; Thoits, 1992; Stryker & Serpe, 1994; Reitzes & Mutran, 1994). Identity prominence is the worth or value individuals place on particular identities (see: Brenner, Serpe, & Stryker, 2014). These past studies justify an experiment of identity prominence since: ratings and rankings are not interchangeable measures of importance, prominence is a measure of identity importance, and prominence hierarchies may be artificial. This study furthers past survey and social psychological research through testing if rankings and ratings are interchangeable measures of prominent obligatory adult identities.

The Effect of Noncontingent Monetary Incentives on Response Rates of College Students
Mariesa Hawkins, *American Institutes for Research*
Mark Masterton, *American Institutes for Research*

College students have their own set of circumstances and experiences that distinguish them from other sub-populations. Various methods have been used to increase response rates of surveys administered to college students, including noncontingent monetary incentives. This poster will report on efforts to use noncontingent incentives in a longitudinal web survey of college students in Nebraska. The first wave of this survey was administered in fall 2015, and collected baseline student characteristics, traits, attitudes, and behaviors. Approximately 1,500 students from three campuses of the University of Nebraska were selected to participate. Approximately 800 of these students were part of an intervention designed to improve their success in college, while approximately 700 were part of a control group. For this survey, the noncontingent monetary incentives were Amazon gift cards in denominations of $20.00 or $35.00. The amount of the incentive depended on whether the student was in the treatment or control group. Control group students were expected to be less inclined to

participate in the survey and were thus offered a larger monetary incentive. This poster will explore the relationship between the redemption of noncontingent monetary incentives and response rates in a college student population. We will examine the response rates of students who redeemed the noncontingent monetary incentive and those who did not. We will also explore any underlying patterns in the redemption of incentives within the population. Results from this research will impact decisions about how to use Noncontingent monetary incentives.

Can Administrative Data Compensate for Cell Phone Under-representation?
David Shor, *Civis Analytics*

As landline surveys become less representative and cell phones become more expensive, survey researchers face a hard choice between bias and variance. Increasing the proportion of cell phone respondents can increase representation, however, this often comes at the cost of reduced overall power for their study. Additionally it is unclear where and how online survey research should be used as an alternative. While others have investigated ways to address this challenge using a variety of new techniques (Barber, Mann, et al., 2014, Levine and Harter, 2009, Chang and Krosnick, 2008, Stewart 2006) an understudied question is if these challenges can be addressed by leveraging administrative data sources such as voter registration records to make samples more representative in a way that is tailored to survey design. To address this gap in the literature this study will investigate the extent to which administrative data can overcome representation issues from under-sampling cellphone respondents and replacing them with respondents from more cost effective modes. This study looks at this issue by examining self-reported partisanship in parallel tracking surveys sampled from voter records where interviews were conducted on landline telephones, cell phones, and online non-probability panels.

A Multi-phase Exploration Into Web-based Panel Respondents: Assessing Differences in Recruitment, Respondents and Responses Between RDD and Web-based Sample Frames
David Redlawsk, *Rutgers University*
Kathleen Rogers, *Rutgers University*
Debbie Borie Holtz, *Rutgers University*

As phone recruitment for surveys continues to be challenging, and the demand for empirical studies assessing differences between randomly recruited phone respondents versus web-based respondents remains high, the Eagleton Center for Public Interest Polling has initiated a multi-phase comparison study to address these questions. ECPIP, established in 1971, is the oldest university-based state survey research center in the United States and regularly conducts a statewide survey (the Rutgers-Eagleton Poll) of the New Jersey adult population on a wide variety of statewide and national issues. Typical recruitment is between 800 and 1000 respondents per survey. Starting with our April 2015 poll, we began to construct a statewide panel by asking respondents if they would be willing to be contacted again in the future for other surveys. In particular, we were interested in obtaining email addresses so that we could add them to a panel for online versions of future Rutgers-Eagleton polls. Our approach has been to field an identical survey directed to respondents recruited from a randomly selected land-based and cell phone sample frame to the web-based panelists. To the maximum extent feasible, the surveys are fielded concurrently. Though the web survey is fielded a bit longer and includes three calls to participate, the survey is closed before we release public information from the phone version to avoid validity threats due to contamination. As the N size of panel respondents increases with

additional recruitment, we expect to randomly select participants from the web-based panel so as to decrease respondent fatigue and related non-response. We will present our findings thus far, assessing the characteristics of our panel – including our recruitment methods, continual solicitation strategies, and attrition – as well as respondent differences between the phone and web-based sample frames and samples themselves, both in terms of demographics and response choices.

Profile of a Problem Gambler: Identifying Gambling Behaviors Across Survey Modes

Janet Cuanas, *NORC at the University of Chicago*
Erin Fordyce, *NORC at the University of Chicago*
Rachel Volberg, *University of Massachusetts Amherst*
Robert Williams, *University of Lethbridge*

The Expanded Gaming Act was signed into law in 2011 allowing Massachusetts to introduce casinos and slot parlors for the first time. As a condition of this legislation, the Massachusetts Gaming Commission was required to conduct research to understand the effects of expanded gaming and to minimize any harmful effects of gambling. The University of Massachusetts-Amherst and NORC partnered to conduct a survey of the general population in Massachusetts prior to the introduction of gambling. The Social and Economic Impacts of Gambling in Massachusetts (SEIGMA) survey inquired about several topics including health status, gambling attitudes and motivation, past year gambling behavior, and gambling problems during 2013 and 2014. This study represents a unique opportunity to survey residents prior to the introduction of gambling, allowing researchers to assess long term impacts. SEIGMA was a multi-mode address-based survey (n=9,578). A primary concern for researchers is the accuracy of self-reported data. The survey asked about sensitive information such as depression and suicide, gambling problems, and drug/alcohol use. Our paper analyzes the impact on reporting sensitive gambling behaviors by mode—phone, web, and paper-based. For example, we discuss the frequency at which "problem gamblers" were identified by each mode. A problem gambler was defined using the Canadian Problem Gambling Index (CPGI), a measure which has been used for other gambling surveys. We also discuss the potential for mode effects on various gambling behavior questions. What mode is more likely to elicit a response and in what question format? In order to build on current literature, we control for known demographic differences in these reported behaviors to determine the magnitude of the self-administered versus interview effect specific to problem gambling.

Increasing Data Flexibility While Maintaining Confidentiality: Tools in Development at the Census Bureau

Amy Lauger, *U. S. Census Bureau*
Michael Freiman, *U. S. Census Bureau*

Data users in government, private industry, non-profits and academia have substantial demand for data from the Census Bureau's surveys and censuses. Hence, the Census Bureau aims to disseminate data widely and with as much detail as possible while keeping the pledge of confidentiality given to all respondents. Although the Census Bureau produces many estimates and tables, along with public use microdata for some surveys, existing data products do not always meet the needs of users. This presentation describes new tools the Census Bureau is developing to provide users with increased flexibility.

One such tool is the Microdata Analysis System (MAS), an online remote access system that was initially intended to allow users to query the official microdata to obtain customized tables and estimates. We

describe unfavorable test results about the data utility and disclosure risk of such a system and explain the resulting change in plans. The Census Bureau now plans to work on two separate tools. The first will be a table aggregator that will allow users much more flexibility for aggregating already published estimates from the American Community Survey. The second will be a synthesized or perturbed American Community Survey microdata file, which will either be directly provided to users or used as a source in a redesigned Microdata Analysis System (MAS).

The Effect of Emphasizing the Web Option in a Mixed-mode Establishment Survey

Jennifer E. O'Brien, *Westat*
Sushama Rajapaksa, *Westat*
Brenda Schafer, *Internal Revenue Service*
Patrick Langetieg, *Internal Revenue Service*

In a series of mixed-mode establishment surveys fielded for the IRS since 2010, we have noticed higher response rates via the web survey option than the mail survey. Moreover, some respondents have reported that that wished they had noticed the web option before completing the mail survey because they would have preferred to complete the survey online. In light of these two observations, we test the impact of making the web option more visible. Some researchers have observed a negative effect of concurrent mode presentation on overall response rate (Medway & Fulton, 2012). On the other hand, there is literature that suggests that giving respondents the option to respond via their preferred mode increases overall response rates (Olson, Smyth, & Wood, 2012). Further, very few of these examinations have investigated these effects in the administration of an establishment survey. In the present study, we changed the location of web access information in a test version of the paper version of the Tax Year 2014 Information Return Burden Survey and measured the effect on overall response rate as well as the number of web vs. mail survey completes. Our two hypotheses were: 1) Making the web option more prevalent in one experimental condition will increase the overall response rate due to the higher visibility of the preferred mode rather than depressing response rates as predicted by the Paradox of Choice (Schwartz, 2004), and 2) The more prevalently-displayed web option information will result in an increased number of web completes relative to mail completes due to increased awareness of the availability of the web option.

May 14, 2016
Concurrent Session I

Non-probability Samples -- Biases and Estimation

Panel orgnizer: Donsig Jang, *Mathematica Policy Research*

Probability-based sample surveys have encountered critical challenges: frame undercoverage, high nonresponse rates escalating data collection costs, pressure to promptly disseminate high quality data, etc. In contrast, disruptive technology innovation utilizes quick and inexpensive means to collect data: opt-in web surveys. In particular, non-government surveys (i.e., marketing surveys or opinion polls) have rapidly taken advantage of such technology development. With non-probability samples growing in use, AAPOR published a report (2013) attempting to address some central issues. While the importance of methods throughout the entire non-probability data collection process is recognized, the question of

how non-probability sample can help estimate population characteristics is still unclear. There have been many empirical studies assessing bias in non-probability sample based estimates by benchmarking them against a gold standard probability sample. Results are inconclusive, as the magnitude of bias depends on the types of variables and post data collection adjustments. In this session, panels will present recent research from their ongoing effort to maintain rigor and unbiasedness in using a non-probability sample for a representative sample. The first presentation will propose a framework for understanding differences between probability and non-probability sample estimates as a function of four separate components, each with different root causes and requiring different remedial approaches. The second presentation will introduce a "fac-sample" by matching non-probability sample cases to a comparable probability sample and present a way to estimate a standard error and a confidence interval for the population of interest from the appropriately weighted fac-sample. The third presentation will focus on the reduction in estimated bias for propensity score adjustments (PSA) and estimated-control (EC) calibration techniques through an empirical study. The last two presentations will present methods to supplement a probability sample with a non-probability sample. This session will continue to stimulate research on scientific, statistically reliable use of non-probability samples.

Decomposing Selection Effects in Non-probability Samples

Andrew Mercer, *Pew Research Center*
Scott Keeter, *Pew Research Center*
Frauke Kreuter, *University of Maryland, JPSM*

Prior studies have found that survey estimates obtained using non-probability samples are sometimes very close to estimates that use probability-based methods, while at other times they differ substantially. To date, the conditions under which non-probability methods yield comparable estimates remain poorly understood. We propose a framework for understanding differences between probability and non-probability sample estimates as a function of four separate components, each with different root causes and requiring different approaches to remedying. First, differences may be due to confounding, or the presence of unobserved factors associated with both the survey outcome and inclusion in a sample. Second, differences may be due to the absence of certain classes of respondent from non-probability samples (such as the non-internet population in web panels). Third, differences may be due to the over or underrepresentation of certain types of respondents. Finally, differences may be reduced or magnified by post-survey adjustment such as weighting or sample matching. We use machine learning and causal inference methods to evaluate the relative contribution of each of these components in explaining observed differences between estimates generated from 11 parallel, non-probability web surveys and the Pew Research Center's probability-based American Trends Panel. We look at the degree to which these components are consistent across different sample providers over a wide range of variables. Because the point of comparison is itself a survey rather than true values, these comparisons cannot be said to measure selection bias; however they can shed a great deal of light on dynamics that produce similarities or differences between different non-probability sample providers relative to a common, probability-based point of reference.

Internet-use Propensity for Matching Probability and Non-probability Samples: the "Fac-sample"

Charles DiSogra, *Abt SRBI*

Large, on-line Internet panels are a popular and inexpensive source for survey samples. The opt-in element for joining these panels makes them a pool of convenience for finding respondents. With no sampling frame and thus no known selection probabilities for sample units, samples drawn or

constructed from these panels bypass the standard metrics that would be applied to a probability sample based on traditional sampling theory. Specifically, a generalizable population estimate standard error and the computed confidence interval for that estimate are missing. Once a large non-probability survey is completed, there may be a post-survey opportunity for the researcher to match these cases to an existing comparable probability sample to arrive at a facsimile sample or "fac-sample" that might be theorized as one of the infinite number of possible probability samples drawn from a population. Because opt-in panels are exclusively Internet users and a true probability sample from the general population would have non-Internet users in it, a logistic model utilizing an array of variables common to both samples can generate Internet non-user propensity scores from the combined samples. A fac-sample can be assembled using this propensity score for matching the non-probability cases to the referent probability sample. With this accomplished, a standard error, a traditional margin of error and a confidence interval can be estimated for the population of interest from the appropriately weighted fac-sample. Inherent bias in the opt-in panel cases may not be entirely addressed or removed by either the matching or population weighting, but it may be quantified and thus known by a comparison with the estimate derived from the referent probability sample. Understanding the inherent bias may help in calibrating or at least interpreting future similar non-probability sample estimates.

Estimated-control Calibrated Estimates from Nonprobability Surveys
Jill A. Dever, *RTI International*

Nonprobability (or design-free) surveys are becoming more common place based on changes in the definition of fit for purpose for many studies. Nonprobability surveys are touted as offering both increased speed in obtaining data on emerging issues (e.g., an opt-in web survey) and decreased costs compared with probability-based surveys. However, evaluation studies have shown that many nonprobability estimates are biased because of errors associated with coverage, selection, and model misspecification. Calibrating design-based survey weights to control totals estimated from other surveys has been implemented for years. Referred to as estimated-control (EC) calibration, this technique has been shown to reduce bias for design-based estimates beyond levels seen when calibrating to typical controls alone such as demographic characteristics. By comparison, propensity score adjustments (PSA) are used to calculate estimates from nonprobability surveys, and may include questionnaire items (e.g., attitudinal questions) as model covariates. However, research to date on PSA shows mixed results. This research builds on prior results related to weighting for nonprobability surveys. We compare the reduction in estimated bias for PSA and EC calibration techniques through an empirical study, noting that the latter methodology has many benefits. Applications are applied where appropriate to a nonprobability survey of young adults enrolled in a smoking cessation program. Next, we outline steps in calculating precision for the estimates when using EC calibration. The discussion concludes with out next steps in the study of nonprobability surveys.

Efficient Estimation using Purposive Supplements to Probability Samples
Avi Singh, *NORC at the University of Chicago*

With the growing demand for Fit-For-Purpose Surveys to save cost by not using rigorous data collection methods, old issues of whether suitable inferences can be made from purposive or nonprobability samples are again at the forefront. For purposive samples, design-based methods are clearly not suitable. There is, however, the possibility of using model-based methods but they assume the design to be non-informative; i.e., the model is assumed to hold for the sample so that an asymptotically optimal prediction of the unseen total from the seen can be made under the model. In addition to the above concern, a second concern with any model-based method is that it is subject to potential

misspecification of the model mean resulting in bias even if the design is non-informative. To overcome these concerns, an alternative approach termed Model-Over-Design (MOD)-Integration for a simplified problem is proposed under the joint design-model randomization where the availability of a core probability sample for which the purposive sample acts as a supplement is assumed. A design-based estimate such as GREG is first constructed which uses a correction for model-error to the mean predictor (i.e., the synthetic estimator). Next, the above model-error correction is improved by using an estimator from the seen in the purposive sample. We remark that while the initial probability sample is used for both estimation of model parameters to obtain a synthetic estimator and for estimation of model-error, the purposive supplement is only used to improve the model-error correction from the additional seen units. The mean squared error (MSE) of the resulting estimator can be estimated using Taylor linearization under the joint randomization of the nonprobability sample (viewed as a nature-made probability sample but unknown), probability sample (man-made and hence known), and the model for the finite population.

Exploration of Methods for Blending Unconventional Samples and Probability Samples

Hanzhi Zhou, *Mathematica Policy Research*
Michael Sinclair, *Mathematica Policy Research*
Jonathan Gellar, *Mathematica Policy Research*

With increasing survey costs, reduced participation rates, and a greater need for timely and domain specific estimates to inform treatment and policy decisions, interests has grown in the use of data sources obtained from unconventional sampling methods (Wang et al., 2014) to supplement or replace traditional survey programs. Given these unconventional data sources may have an unknown, and likewise uncorrectable levels of bias in their estimates, we explore the use of composite model-based estimation methods in conjunction with a Bayesian procedure for variance estimation to blend data from an unconventional sample with parallel data from a traditional probability-based sampling strategy. With proper assumptions on the relative cost and bias for the two data sources, we study the utility of this blending approach to adopt at least a partial traditional survey replacement action plan in two primary settings: (1) augment a traditional probability sample survey with a faster-turnaround and more economy unconventional sample component; and (2) use a smaller traditional probability sample survey to validate or correct the unconventional survey data for the observed biases. We present a simulation study based on public use data from the National Health Interview Survey (NHIS). We compare the estimates obtained from the unconventional and traditional sources individually, and in combination under various scenarios with respect to the sample sizes, the levels of bias in the unconventional sample, along with the use of calibration and model-based estimation procedures to correct for these biases using two sets of available covariates. Through cost-efficiency analysis across scenarios, we show that the use of the proposed methods enables one to determine the fitness of use of an unconventional data source and to leverage such data appropriately to fulfill program objectives.

Identifying Respondents Behaving Badly

What Would You Ask?: Exploring Why Interviewers Select Different Techniques to Reduce Question Sensitivity

Robin L. Kaplan, *U.S. Bureau of Labor Statistics*
Erica C. YuWright, *U.S. Bureau of Labor Statistics*

Interviewers often face the challenge of asking respondents to reveal potentially sensitive information about themselves without knowing beforehand how the respondent will react. To reduce question sensitivity, sometimes interviewers use forgiving wording (i.e., positively loading questions to invite a socially undesirable answer). Other times interviewers use techniques such as distancing themselves from the survey (e.g., "I have to ask every question as worded") or apologizing (e.g., "I'm sorry to ask this..."). Little empirical evidence exists on the effectiveness of these techniques, why interviewers select one technique over another, and respondents' reactions to them. In this experiment, 450 participants completed an online task where they were instructed to take on the role of survey interviewer in a series of online interactions with a respondent. Participants were randomly assigned to one of two survey goal conditions: (a) elicit honest and substantive responses from the respondent, or (b) no specific survey goals. Half of the survey interactions were placed in a sensitive context (e.g., a respondent struggling to find work due to age discrimination) and half were placed in a neutral context (e.g., a respondent who is voluntarily retired). Participants then selected the question technique (forgiving wording, distancing, apologizing, or a direct question) they most preferred to ask each respondent in the given context. Afterward, participants indicated why they made each wording selection (to increase honest reporting, to reduce question sensitivity, to make the question more comfortable to ask, or some other reason), along with sensitivity ratings for each type of question. We compare wording technique selection, reasons for wording technique selection, and sensitivity ratings of the questions across each survey context, along with individual difference factors that may impact wording selection. We discuss the implications for when and why interviewers might use different sensitivity-reduction techniques and their effectiveness across survey contexts.

Habitual and Non-habitual Respondents in a Longitudinal Telephone Study: Differences and Implications

Eran N. Ben-Porath, *SSRS*
Bianca DiJulio, *The Henry J. Kaiser Family Foundation*
Mollyann Brodie, *The Henry J. Kaiser Family Foundation*
Jamie Firth, *The Henry J. Kaiser Family Foundation*

Longitudinal studies are an important tool to observe and better understand changes in attitudes and behaviors over time, both across a population and among individuals themselves. While the ability to track changes in individuals over time is what makes longitudinal studies unique and insightful, these types of studies are also prone to systematic differences between those respondents who participate in all waves of the study and those who participate less consistently. This differentiated pattern of response introduces the risk of bias if habitual respondents (those responding to all study waves) differ from the less cooperative ones (non-habitual respondents) in terms of key characteristics, and even more so, if their responses differ on the main attitudes or behaviors being measured. Looking at data from a multi-wave longitudinal RDD telephone-based study of uninsured California residents, this paper focuses on the differences between habitual and non-habitual respondents, focusing on a wide range of demographics and reported behaviors and attitudes surrounding health insurance coverage and health

care. The analysis indicates that the propensity to respond consistently to all waves of the study is associated with some demographics such as language-spoken, age and education. On most of the central attitudes and behaviors measured, some small differences were observed, however most of these differences were reduced-further by weighting (albeit with some loss of statistical power). While the findings show a consistent systematic differentiation between those respondents who continue to participate and the more reluctant respondents, as of the first three waves, the impact on the survey results is minimal. Attrition results from the fourth and final wave, to be completed in the coming months, will also be included. In our discussion we provide guidelines for measuring the impact of differences between habitual and non-habitual respondents, and ways to minimize this impact.

The Three Dimensions of Distracted Driving: Structural Equation Modeling of Driving Behavior

Stanislav Kolenikov, *Abt SRBI*
Paul Schroeder, *Abt SRBI*
Melanie Wilbur, *Abt SRBI*
Kristie Johnson, *NHTSA*

The 2015 NHTSA Survey of Distracted Driving Attitudes and Behavior seeks to measure the perceptions and driving habits of drivers 16 years of age or older in the US. A total of 6,001 interviews which were conducted among a national representative sample of individuals 16 or older who had driven a motor vehicle. To account for the current shift to cell phone use and the underrepresentation of younger individuals in samples using landline telephones, a partial overlapping dual sampling frame of households with landline phones, and households that relied only or mostly on cell phones, together with a landline phone oversample of individuals ages 16 to 34 was used. We utilize a popular multivariate statistical technique of structural equation modeling to link the common factors of distracted driving to the outcomes on the road. We identified three dimensions of distracted driving: talking to other people in the vehicle; using smartphone; and addressing personal needs, which include eating, drinking, and controlling the car audio. Structural equation modeling shows that using a phone is associated with being stopped by the police, while addressing personal needs is associated with higher likelihood of crashes or near crashes.

User Experience and Eye-tracking: Results to Optimize Completion of a Web Survey and Website Design

Lauren A. Walton, *Nielsen*
Kay Ricci, *Nielsen*
Amanda Libman Barry, *Nielsen*
Christina Eiginger, *Nielsen*
Leah Christian, *Nielsen*

Asking respondents to sign up for online research panel is a difficult task. It requires requesting multiple pieces of information, such as demographics, devices presence/usage details, and sometimes to install an app on their device. However, following best practices in website design and visual design principles can help respondents navigate an online registration survey. This study focuses on using eye-tracking, a technology that allows us to passively study people's eye movements, and other usability measures to analyze how respondents perceive, understand and experience different designs of the online panel registration surveys (n=16). Eye tracking has been used, mostly for questionnaire testing, within the the survey research field (see Redline and Lankford 2001, Graesser et al. 2006, Galesic et al. 2008, and

Lenzer, Kaczmirek, and Galesic 2011). In this study, eye tracking is used to gain insights into the perceptions of four online panel registration websites each with a unique layout, messaging, and visuals. For two of these websites, participants completed the registration survey while eye tracking data was also collected. This paper centers on the collection and analysis of qualitative and quantitative measures of the user experience, including eye tracking data (e.g., fixation duration and time to first fixation), participants' verbalizations, and self-reported satisfaction from in-depth debrief interviews. Results include that color and proximity of the "join now" button can drive participants to look or not look at it. Participants preferred images that conveyed the task they were being asked to do versus happy smiling people or icons. The paper will provide recommendations about the design of online registrations surveys utilizing gaze plots, heat maps, and qualitative debrief data, to inform the recommendations.

Straightlining: Who Does It and When?
Florian Keusch, *University of Mannheim*
Ting Yan, *Westat*

Non-differentiation in grid questions, also known as straightlining, happens when respondents select the same response option on all items instead of carefully answering each item individually. Although identical responses to a set of related items worded in the same direction is not necessarily a sign of careless answering, straightlining on items from different domains or on both positively and negatively worded items is considered a sign of low cognitive effort. Previous research has shown that straightlining correlates highly with other satisficing behaviors (such as speeding) and therefore is an indicator of poor data quality. Especially in online panels, where respondents might be repeatedly exposed to grid questions, straightlining is a potential issue. Earlier research on straightlining gives emphasis on characteristics of respondents who are more likely to straightline. However, there is not enough attention paid to question characteristics that potentially attract straightlining. This study considers straightlining as the result of an interaction between respondent and question characteristics. Using data from the LISS online panel in the Netherlands, we conducted multilevel modeling to identify who are more likely to straightline on what kinds of survey questions.

Matchmaking Survey Data and Big Data

Integrating Big Data and Panel-based Estimates: A Practical Application from Audience Measurement
Christine Pierce, *Nielsen*
Diane Morovati, *Nielsen*

The digitization of media content has brought both challenges and benefits to the field of audience measurement. The proliferation of content across multiple platforms has led to increasingly fragmented audiences, which has, in turn, constrained the ability to estimate these audiences using surveys or panels alone. At the same time, digitization enables the creation of digital tracking data which provides "Big Data" for measurement. While the integration of Big Data provides great promise, it also presents new challenges that must be mitigated. This paper describes the measurement of digital content ratings and the transition from a panel-centric approach to one that incorporates both Big Data and panels. The digital data is generated passively from devices rather than collected actively from people who have agreed to be measured. As such, it provides greater stability in the estimation of volume, e.g., how many views. However, it does not provide the requisite demographic characteristics, e.g., who viewed. In addition, there are coverage gaps as some content providers do not deploy the digital tags that enable

measurement. In order to address the biases inherent with the digital tracking data alone, Nielsen developed methods to combine digitally generated data with representative data collected in its electronically measured online panel. This combined approach leverages the benefits of Big Data and panel data to achieve the highest possible information to data ratio (Groves 2011). Comparisons of the Big Data and the panel-based estimates will be shared, along with the corresponding estimates of variance and bias, highlighting the benefits and drawbacks of each data source. Lastly, the paper demonstrates practical approaches for combining these data sources in order to provide comprehensive audience measurement solutions.

The Stability of Economic Correlations Over Time: Comparing Data from Gallup's Daily Tracking Poll, Michigan's Surveys of Consumers, the S&P 500 and Twitter

Josh Pasek, *University of Michigan*
H. Yanna Yan, *University of Michigan*
Frederick G. Conrad, *University of Michigan*
Frank Newport, *The Gallup Organization*
Stephanie Marken, *The Gallup Organization*

Economic confidence measures have long been viewed as important correlates of stock market performance and future economic growth. A recent study by O'Connor and colleagues suggested that it might be possible to tap economic confidence by tracking the sentiment associated with the keyword "jobs" on Twitter. If effective, this method could reduce the cost of measuring perceptions of the economy compared to more traditional survey approaches, or provide a method of augmenting those traditional measures with additional information. The current study examines the consistency of relations between the Twitter data (i.e. the sentiment of tweets containing the word "jobs") and survey measures. To assess the reliability of the Twitter metric, we compare Twitter sentiment around the keyword "jobs" derived using Lexicoder to evaluations of the economy based on responses to a question from Gallup's daily national cross-sectional survey on whether the economy is currently getting better or worse and a similar question from the University of Michigan's Surveys of Consumers about whether the economy has improved or worsened over the prior year, and finally to concurrent data from the S&P500 index. Comparing these different data sources reveals that similarities between all four datasets vary over time, from nearly zero correlations at some moments to correlations of greater than .8. In particular, we find that there is a set of identifiable moments when all data sources appear to trend together as well as a number of periods when the data sources are unrelated. We attribute these results to the influence of large-scale events on both the economy and economic confidence. The findings yield the conclusion that Twitter data and survey data do not necessarily capture the same stories about daily happenings, but will frequently do so over larger periods.

Big Survey Data + Big Digital Data + Big Media Data

Tony Foleno, *The Ad Council*
Sheri Klein, *The Ad Council*
Chris Jackson, *Ipsos Public Affairs*

The Ad Council runs 40+ public service communications campaigns. As part of campaign evaluation procedures, we conduct national tracking surveys that gauge relevant awareness, attitudinal and behavioral measures over time. In autumn 2014, we initiated a large continuous omnibus tracking survey with Ipsos Public Affairs. The survey is never out of field. To date, we have collected more than 54,000 interviews concerning 25 different social and health issues across 25 different national target

populations. In addition to this survey dataset, we maintain other massive datasets to aid in evaluative research, including media data (platform, impressions, dollar value) and digital engagement data (website analytics, social media metrics). Our task at hand is to integrate these disparate data streams into a cohesive narrative that assesses what is working in a given media campaign, what is not, and what to do about it. These analyses go well beyond simply correlating survey trends with other data trends. In this presentation, we will present an overview of the Children's Oral Health campaign, demonstrating how we integrated a weekly survey dataset from 2014-2015 with other data. We will also present data relating to our campaign with Goodwill, demonstrating how survey insights and a media mix analysis demonstrated that the campaign drove more than 360 million pounds of goods donations to Goodwill. The Ad Council campaign is at least partially responsible for these shifts in behaviors. We argue that in today's complex, fragmented media and marketing landscape, survey measurement—no matter how robust—may fall short of telling the full story of a media campaign if it's not integrated thoughtfully into other data trends.

Using Administrative Records in the American Community Survey

Amy B. O'Hara, *U.S. Census Bureau*
Deborah Stempowski, *U.S. Census Bureau*

The Census Bureau is looking for ways to reduce the difficulty and length of the American Community Survey (ACS) using administrative records. We have identified sources of federal, state, and commercial data that may alleviate the need to ask certain questions altogether or for a subset of the ACS sample. Work is underway to acquire additional data sources and assess the coverage of ACS persons and households. This paper describes analyses of ACS questions that are candidates for external data enhancement or replacement. These include housing variables such as year built, number of bedrooms, home value, and property taxes, and person questions including residence one year ago and the multi-part income question. For each topic, we discuss whether the ACS question is sensitive or cognitively difficult, and summarize the quality characteristics of available administrative records data. We discuss matching challenges and results. Based on these initial findings for coverage and concept alignment, teams of ACS specialists are determining whether each question is a good candidate for removal. Recommendations will be based on an assessment of the implications of implementing such a change, considering data quality, reliability, alignment of reference periods, and the impact of a break in series. Our research teams are consulting with methodologists and stakeholders regarding the appropriateness of direct substitution and the impact on data availability and published estimates.

Rights of Respondents

Examining the Use of Privacy Language: Privacy from the Respondent's View

Kay Ricci, *Nielsen*
Lauren A. Walton, *Nielsen*
Ally Glerum, *Nielsen*
Robin Gentry, *Nielsen*

As the technological landscape has evolved over the last decade, privacy concerns have become a prevalent headline in the media. As survey researchers, we need to understand how these concerns affect our respondents. We recently re-examined the privacy language on a screener questionnaire, which is a self-administered survey that asks respondents for both demographic information as well as a contact number. In revisiting the language used on the screener, we were interested in not just

exploring the concept of privacy but also the implications and interpretations of new TCPA rules and regulations around automatic telephone dialing of cell phones. This research study explored the impact of this revised privacy language. First, we conducted 46 cognitive interviews in both English and Spanish to assess how respondents perceive privacy and interpret specific concepts such as an "automatic telephone dialing system." These findings were then used to develop the new privacy language that was included in a large-scale impact analysis of the screener questionnaire for Nielsen's TV diary measurement. This survey used a national address based frame, with approximately 860,000 households receiving the updated language. We evaluated key recruitment metrics such as return rates, the demographic distribution of cooperating households, and the number of households that provided a contact number compared to prior data collection cycles. This paper will provide key insights into a respondent's perspective on privacy language as well as how privacy language affected participation in the screener survey.

Allowing Access to Household Internet Traffic: Maximizing Acceptance of Internet Measurement
Megan Sever, *Nielsen*
Sean Calvert, *Nielsen*

Household internet penetration continues to grow in the US – indeed, the percentage of American adults who use the internet has increased from 52% in 2000 to 84% in 2015[1]. As such, members of the public are increasingly accessing information, consuming media, and purchasing products using the internet in their own home using a variety of devices. To keep up with this trend, research organizations are continually developing methodologies for measuring what people watch and buy online in their home. While some of these methodologies utilize data collected by third parties (such as Amazon, Google, Facebook and Twitter), they can lack complete coverage and accurate audience demographic information (to know who was watching or buying), limiting the usefulness of their data. Alternatively, panel-based methodologies, which collect household internet traffic, provide both complete coverage of the internet traffic in the home and individual demographic information. For panel-based methodologies to yield representative data, procedures need to be developed to maximize acceptance rates in order to minimize non-response bias. To evaluate acceptance of household internet data collection, we conducted a series of in-depth interviews in which participants were mock-recruited to participate in a measurement requiring access to their household internet traffic and Wi-Fi network by providing their Wi-Fi password. This approach afforded the ability to identify potential participants' areas of concern and questions around household internet traffic collection and explore various strategies to encourage participation in this type of measurement. This paper discusses the key findings from the qualitative research including 1) Factors leading to acceptance, 2) Impact of sponsorship or reputation of the research organization making the request, and 3) Information provided that can help allay privacy and security concerns to increase acceptance [1] Andrew Perrin, Maeve Duggan, "Americans' Internet Access: 2000-2015." Pew Research Center, June 2015

Informed Consent: What Do Respondents Want to Know Before Survey Participation?

Nicole R. Buttermore, *GfK Custom Research*
Randall K. Thomas, *GfK Custom Research*
Jordon Peugh, *SSRS*
Frances M. Barlas, *GfK Custom Research*
Mansour Fahimi, *GfK Custom Research*

In 2015, proposed changes to the AAPOR Code of Professional Ethics and Practices regarding disclosure of survey sponsorship led to a vigorous debate within the AAPOR community. Central to this debate is the question of what constitutes informed consent in survey research. While many researchers have considered what, in their opinion, constitutes informed consent, few have actually asked survey respondents what information they deem important to make an informed decision about participation. We conducted a survey using a probability-based sample from an online panel and asked members about the level of risk they perceive when completing a survey and the type of information they feel researchers should disclose. Although the majority of respondents reported that taking online surveys constitutes little or no risk to a person's mental health, about a quarter of respondents indicated they had stopped taking an online survey in the past year because it made them feel uncomfortable. Among other questions, respondents were asked to rate the importance of receiving information about the organization conducting the research, the organization paying for the research, protection of personal information, survey topic, survey risks, incentives, and interview duration. The only piece of information rated as essential by the majority of respondents was about the protection of personal information and privacy provisions. We will compare results from the online panel sample to those from a dual-frame RDD telephone survey. Results will be discussed in the context of better understanding important elements of informed consent for surveys conducted in different modes and as part of participating in a panel compared to a one time survey request.

Communicating Data Use and Privacy: In-person versus Web based methods for message testing

Aleia Clark Fobia, *U.S. Census Bureau*
Jennifer Hunter Childs, *U.S. Census Bureau*

Communicating messages about privacy, data use and access, and confidentiality is critical to earning and keeping the trust of respondents, and to ensure their willingness to participate in surveys. In addition, informing respondents about their rights and how their data will be used is often required by law. In this paper, we present the results of both Internet and in-person methods for testing respondent comprehension of messages that convey legally required information about privacy and confidentiality. Using an open-ended Internet instrument, we collect qualitative data on respondent comprehension of privacy and data use messages. Analysis of this data identifies high and low performing messages to be included in a smaller-scale cognitive test. Combining these two methods allows for a larger scale data collection than usual in a qualitative study, while retaining the ability to elicit rich description and spontaneous probing. We explore the advantages and limitations of using open-ended questions on the web for testing survey questions. Through direct comparison with more traditional methods, this research will contribute to literature on alternatives to traditional in-person cognitive interviewing.

Respondent Burden & the Impact of Respondent Interest, Item Sensitivity and Perceived Length

Morgan Earp, *U.S. Bureau of Labor Statistics*
Erica C. YuWright, *U.S. Bureau of Labor Statistics*

Using previous research on survey respondent burden (Chapin, 1920; Bradburn, 1978; Sharp and Frankel, 1983; Fricker, Yan, & Tsai, 2014), we developed an index of burden including and exploring the effects of respondent interest, item sensitivity, and perceived length. This index was administered at the end of an online survey via Amazon Mechanical Turk. The survey provided participants with differing "burden" experiences in order to understand the underlying factors of respondent burden and how to best measure it. The differing experiences were created through control of the survey content, which varied between groups on dimensions of burden that have been found consistently in previous research: interest, difficulty, and sensitivity. Survey questions that varied along these factors were selected from several current federal surveys. The overall index was first assessed in terms of reliability, measurement error, and differential item functioning using Item Response Theory. Using items with good fit, we then assessed the results of the experiment to determine if the factors of interest, sensitivity, difficulty, and perceived length affected the overall respondent burden of participants and the probability of responding in a follow-up survey.

God and Hillary, not necessarily in that order

Are We Ready for a Woman President? Priming Experiments with Gender and Candidate Preference

Krista E. Jenkins, *Fairleigh Dickinson University*
Dan Cassino, *Fairleigh Dickinson University*
Peter Woolley, *Fairleigh Dickinson University*

This current research in the 2016 presidential electoral campaigns replicates and modifies survey experiments done in the 2008 Democratic presidential primaries to explore the effect of the candidate's gender on voters' preferences. Many public opinion measurements purport to have shown a trend toward greater willingness of the electorate to support a female president (Kohut 2007), but priming treatment groups to think about gender suggests otherwise. For example, in 2008 we found that when voters were primed to think about gender using a variety of survey questions concerning spousal pay and women's share of the electorate, among other things, voter support for Hillary Clinton was reduced in the Democratic primaries. Although the data did not indicate that being female categorically hurt Senator Clinton, it did suggest that the relationship between gender and public support for her candidacy was complex. Fast forward to recent months and our national and statewide surveys have again found a similar dynamic. That is, when voters, especially male voters, are presented with questions about gender inequality, we are finding that Clinton is more hurt than helped as compared with those who are asked about her candidacy before being asked about gender inequality. In short, this research, based on multiple national and regional surveys, raises the continued specter of gender stereotyping in electoral decision making. Even though society wants to be done with using immutable characteristics like gender and race to define a person, the bottom line may be rather simple: gender stereotyping is alive and well. As difficult as it might be to get respondents to admit their biases, our analyses demonstrate clearly that survey research can help to validate what is often suspected yet rarely admitted. Namely, far from being a moot question for voters, gender matters.

Religious Right 2.0? Religion, Religious Nationalism and the Tea Party Movement

Malaena J. Taylor, *University of Connecticut*
Ruth Braunstein, *University of Connecticut*

Previous research evaluating the relationship between the Tea Party Movement (TPM) and the Religious Right (RR) has yielded mixed results. While some suggests significant overlap between these movements, others highlight the secular nature of the TPM's organizational infrastructure, goals, and culture. Meanwhile, some research suggests that the TPM may represent a new kind of religiously infused movement, distinct from the RR but including some elements that may resonate with its members. We attribute these mixed results to disagreement over how to conceptualize and assess religiosity and how to interpret the religious composition of the TPM as a whole, in light of factions within this movement. Drawing on analysis of a Public Religion Research Institute dataset that allows us to compare TPM and RR participants across measures of religious affiliation, religious orthodoxy, religious commitment, and religious nationalism, this paper aims to resolve some of these contradictory findings. First, a comparison of the TPM and the RR confirms that despite overlapping memberships, there are also important differences in the religiosity of these movements' members. Second, a within-TPM comparison – of those participants who are also part of the RR and those who are not – confirms that differences between the two movements on average are the result of significant religious heterogeneity within the TPM. Namely, while half of TPM members are also part of the RR, those who are not part of RR have very different orientations to religion. At the same time, a sizable majority of Tea Partiers share religious nationalist views, despite these religious differences.

Measuring Religious Identity Among U.S. Protestants

Gregory A. Smith, *Pew Research Center*
Claire Gecewicz, *Pew Research Center*
Elizabeth Sciupac, *Pew Research Center*

Many political and social surveys include questions about religious identity because it is strongly correlated with social and political attitudes as well as political behavior. These surveys often employ a shortcut for measuring religious identity among Protestants; rather than ascertaining respondents' actual denominational affiliation (e.g., Southern Baptist Convention, United Methodist, etc.) and then grouping denominations into traditions (e.g., evangelical Protestantism, mainline Protestantism, the historically black Protestant tradition), they instead simply ask self-identified Protestants whether they think of themselves as "born-again or evangelical Christians." Protestants responding yes to this question are grouped together as "evangelicals," while those answering no are categorized as "mainline Protestants." This paper uses data from the Pew Research Center's 2014 Religious Landscape Study to assess the performance of the shortcut approach for measuring religious identity. While most social and political surveys have room for just a couple of religion questions, the Religious Landscape Survey includes both the "born-again" question and an extensive list of questions about religious identity, beliefs and practices. How well do the evangelical and mainline groups as defined using the "born-again" question map onto actual denominational affiliation? That is, do "evangelicals" as defined using the shortcut approach actually belong to evangelical denominations, and do "mainline Protestants" as defined with the shortcut approach belong to mainline denominations? Beyond religious identity, how do other beliefs and practices correlate with responses to the "born-again" question? We also compare the results from the 2014 study with findings from a similar study we conducted in 2007, to assess how the performance of the shortcut approach may be changing over time. Answering these questions will help researchers better understand what they are measuring when using the "shortcut" approach to religious identification, and thus better understand the links between religion and politics.

Religion, Evolution and Abortion: How Religion Influences Attitudes Towards Social Issues
Sarah Kelley, *University of California Berkeley*

Previous research has identified the importance of different schemas about the relationship between God and humanity in mediating the relationship between religion and attitudes towards abortion. I approach this issue quantitatively by demonstrating that those who interpret the world through a Darwinian schema – seeing humanity as naturally evolving rather than part of a directly controlled divine plan – are substantially more supportive of abortion than those who do not. I use Structural Equation Modelling to untangle the relationships between acceptance of the Darwinian Schema and attitudes about abortion, taking into account also religious upbringing, faith, church attendance, beliefs about life at conception, and attitudes about pre-marital sex. On average, a person who rejects the Darwinian schema will be a full 33 points out of 100 less supportive of abortion than an other wise equal (equally religious, with a similar religious upbringing etc) person who embraces the Darwinian interpretation. The relationship between acceptance of the Darwinian Schema and attitudes towards abortion demonstrates the importance of this previously under-studied schema in understanding social issues. There are clear conceptual links between understanding humanity as naturally evolved rather than divinely created and attitudes about important social issues like abortion, stem cell research, homosexuality and the death penalty (among others) and further research is needed to understand these links. Furthermore, this illustrates the role that advanced data analytic techniques like Structual Equation Modelling can play in exploring complex webs of social attitudes Data comes from the 2009 "Attitudes towards Stem Cell Research 5" survey (N=1680).

The State of Cell Phone Surveying

Implications of Manually Dialing Cell Phone Sample After the FCC Declaratory Ruling
Jonathan Best, *Princeton Survey Research Associates International*
Evans Witt, *Princeton Survey Research Associates International*

In July 2015, the Federal Communications Commission issued a Declaratory Ruling revising its interpretation of the Telephone Consumer Protection Act (TCPA). One aspect of the FCC ruling was a broadening of the definition of an autodialer. As the industry grappled with how to implement the ruling, PSRAI decided to test the impact of manually dialing cell phone sample on a system without the capability of automated dialing, compared to having a human interviewer initiate each call with explicit instructions to a system did not use automated dialing (although it had the capability). Two experiments were run, each with identical questionnaires on two separate cell phone samples, differing only in the dialing method. We analyzed the data to compare sample and interviewer performance, productivity rates and sample dispositions and other metrics.

Propensity Modeling in a Cellular Environment
Paul William Burton, *University of Michigan*

Dual-frame and cell-only phone studies have become a permanent feature of the survey research industry and present considerations that are relatively new to the field. Dutwin and Malarek (2014) show that excluding cell phone numbers with inactive and unknown activity codes can increase

productivity with a minimal reduction in coverage. Marketing Systems Group (MSG) provides detailed activity codes for RDD cell phone samples which indicate categorically the length of time the cell phone number has been active or inactive. Little research exists on the use of activity codes and other sample frame variables to predict the likelihood, prior to data collection, of contact or complete interviews with RDD cell phone numbers in order to target cases and increase operational efficiency. Using the study results from a nationally representative cell phone only RDD sample with a repeated cross sectional design, this investigation builds models that predict the likelihood of contact and completed interviews with varying levels of accuracy. This is done by taking advantage of the MSG's activity codes and appending ACS and Decennial survey data through the billing center county-level FIPS information. This study has put into practice Dutwin and Malarek's (2014) suggestions regarding truncation of activity codes in the sample. At the same time, the potential for bias must be examined, and previous waves of survey data will be used to demonstrate if respondents with high predicted response propensities differ significantly on key statistics and demographics when compared to respondents with lower predicted propensities.

Are Women Less Likely to Answer Cell Phones?

Sarah Dipko, *Westat*
Darby Steiger, *Westat*
Shannan Catalano, *U.S. Bureau of Justice Statistics*

Two small feasibility tests and a large pilot study conducted with RDD cell phone samples in 2014 and 2015 yielded surprisingly low eligibility rates for adult women. These data collections were part of a study funded by the Bureau of Justice Statistics to develop a survey on rape and sexual assault. The study protocol screened for adult females. The assumed eligibility rate was 40 percent, based on estimates of cell phone ownership by minors versus adults and by gender. The actual eligibility rate observed for all three data collections ranged from 23 to 27 percent. The rate of reaching adults was close to expectations (74 to 78 percent of all screened phone numbers, compared to the assumed 80 percent); however among screened adults the rate of reaching women was much lower than expected (just 30 to 36 percent, compared to the expected 50 percent). Results from the feasibility tests (both of which used samples of roughly 1,300 cell phone numbers) were presented at AAPOR 2015, and suggested that women were more likely than men to never answer their cell phones – 78 percent of final non-contacts from the second test for which gender of voice mail could be determined were coded as female; however among final refusals an even gender split was observed. This paper focuses on results from the much-larger pilot study which employed a sample of 105,498 cell phone numbers, to examine the robustness of the earlier findings and also to examine consistency among interviewers in coding gender based on voice of the respondent. In addition, accuracy of gender coding by interviewers for both non-contacts and for refusals will be examined by comparing this coding to final case results for which gender is known (e.g., completed screeners with eligible adult women, and ineligible screeners completed with adult males).

Evaluation of Cell Phone Non-coverage in Large Metropolitan Areas
Seth Brohinsky, *Abt SRBI*
Raphael Nishimura, *Abt SRBI*
Dean Williams, *Abt SRBI*
Scott Clement, *The Washington Post*

For many metropolitan and local-area telephone polls there is a growing non-coverage problem caused by city residents with a cell phone area code outside the geographically defined survey area. Appending billing zip codes and activity flags does not completely correct for this type of non-coverage. In Washington D.C., for instance, it has been estimated that 15 percent of the adult population and 31 percent of cell-only adults are only reachable on cell phone numbers outside of D.C.'s 202 area code. (Skalland, and Khare, 2013). This non-coverage problem in Washington D.C. is not unique; the number of out-of-area cell phone numbers is rising in metropolitan areas across the country. Survey researchers must develop a reliable and cost effective solution to correct for this non-coverage problem.In an attempt to reduce cell phone non-coverage in local-area telephone surveys, this presentation examines two recent polls Abt SRBI has conducted in Philadelphia, PA and Washington D.C. In addition to traditional landline and cellular RDD designs, these polls used supplemental samples of targeted cellular phones with an out-of-city area code. The Philadelphia poll consisted of non-local cell phone numbers with a known residential address in the city. The Washington D.C. poll used a list of registered voters with a known address in the city and a non-202 area code cellphone number. Our analysis measures the accuracy and reliability of each supplemental sample. In addition, we performed analyses with and without the supplemental samples to examine differences in key substantive survey questions, demographic makeup, and the impact on design effects and margin of error. This analysis gives a clear picture of how cellphone under-coverage affects survey results when polling in large metropolitan areas and offers advice on how to best correct for this type of coverage problem.

Assessing the ACA: Measuring Coverage

Adding Complexity to an Already Difficult Task: Monitoring the Impact of the Affordable Care Act (ACA) on the Misreporting of Medicaid Coverage
Brett Fried, *University of Minnesota*
Kathleen Thiede Call, *University of Minnesota*
Elizabeth Lukanen, *University of Minnesota*
Giovann Alarcon, *State Health Access Data Assistance Center*

Research shows that survey respondents experienced difficulty correctly identifying if they were enrolled in Medicaid prior to the implementation of the ACA (Boudreaux et. al. 2015). Specifically, people were better at reporting insurance status (covered or not) than identifying their coverage type (e.g. Medicaid versus Medicare). New coverage options in the ACA have led to gains in health insurance coverage, but have likely also made it more difficult for survey respondents to accurately report coverage type. Reasons include (1) the use of a "no wrong door" policy to determine both marketplace subsidy and Medicaid eligibility (2) the varied implementation of the marketplaces (state-based, federally facilitated or federal-state partnership) and (3) the implementation of the Medicaid expansion through 1115 waivers. Accurate measurement of Medicaid in surveys is critical to understanding population level impacts of the Medicaid expansion.. This study tests the association of state level Medicaid policies with differences between administrative (and survey data). To do this we use regressions to model the ratio of Medicaid/CHIP enrollment as reflected in administrative data from the

Center of Medicare and Medicaid Services (CMS) to enrollment in the American Community Survey and the Current Population Survey by state. Our right hand side policy variables include the type of marketplace, type of Medicaid expansion, Medicaid determination (state or federal) and whether or not it is a Medicaid expansion state. In addition, we control for demographic and economic characteristics. We also compare similar ratios for 2012, 2014 and 2015 and test if these differences are statistically significant. One preliminary finding is that misreporting is higher in states that run their own marketplaces. A potential explanation for this result is that higher levels of media attention and outreach in these marketplaces increased the likelihood that an enrollee misreported their Medicaid enrollment as private coverage.

The Effect of Question Characteristics on Support for the Affordable Care Act (2010-2016)

Tianshu Zhao, *University of Illinois at Chicago*
Timothy P. Johnson, *University of Illinois at Chicago*

Health care reform has continued to be an important issue of debate in national polls. Over the past six years, considerable effort has been invested in measuring and tracking public opinion regarding the Affordable Care Act (ACA). This presentation will investigate trends in American public opinion on support for the ACA from 2010 through early 2016 and attempt to interpret it from a methodological perspective. We utilize data from more than 150 national tracking surveys that continue to examine opinions regarding the ACA, and examine a variety of question-level and survey-level characteristics that preliminary analyses suggest may help us better understand elements of survey methodology that may be associated with support or opposition for the ACA. At the question-level, we examine: complexity of question wording, balance of response options, types of response options, and use of key words and phrases to describe ACA; at the survey-level, we examine: the timing of key historical events related to the ACA, types of survey sample employed, survey house effects, and dates of data collection. Using multi-level modeling, this presentation will address the question of why various national surveys attempting to measure support and/or opposition to the ACA have shown variability in findings.

Validating Self-reported Health Insurance Coverage: Results from an Experiment on CPS and ACS Questions

Joanne Pascale, *U.S. Census Bureau*
Kathleen Thiede Call, *University of Minnesota*
Angela Fertig, *Medica Research Institute*
Don Oellerich, *U.S. Department of Health & Human Services*

Surveys generally derive an estimate of the uninsured by asking about coverage through a range of different sources and then identifying those with no reported source of coverage. Prior to full implementation of the Affordable Care Act (ACA) in 2014, there was ample evidence that survey estimates of insured/uninsured status were reasonably accurate, but that measures of coverage by particular types or sources of coverage (e.g., Medicaid, employer-sponsored insurance (ESI)) were more fraught with measurement error. Recent research focusing on the effects of the ACA indicates that survey reporting of coverage type is even more complicated than it was pre-ACA. To gain a more comprehensive understanding of this measurement error, a large-scale experiment was conducted in which enrollment records from a private health plan were used as sample and randomly assigned to survey treatments - one including the health module from the Current Population Survey (CPS) and one from the American Community Survey (ACS). Records included enrollees in Medicaid, ESI and non-group

coverage both within and outside the Marketplace. Data were collected in the spring of 2015 (n=6,644 individuals). Analysis examines the "absolute" accuracy (records versus survey data), and "relative" accuracy (absolute accuracy of CPS versus ACS). Several correlates of reporting accuracy are explored, including standard demographics, characteristics of the survey reporting task (e.g., household size, relationship between household respondent and other household members), and characteristics of the coverage (e.g., plan type, duration, recency and stability of coverage).

Isolating the Effect of the Affordable Care Act: Tracking the Uninsured
Stephanie Marken, *The Gallup Organization*

In 2008, Gallup and Healthways initiated a partnership to track and understand the key factors that drive well-being. The Gallup-Healthways Well-Being Index provides an in-depth, nearly real-time view of Americans' well-being. The Gallup-Healthways Well-Being Index is a nationally representative survey of adults currently living in the United States interviewed by telephone every day, 350 days per year. Since this study was initiated in 2008, Gallup has measured the insurance status of adults currently living in the United States as part of the Gallup-Healthways Well-Being Index survey. Gallup's longstanding trend on this important metric provides the ability to analyze the insurance rate prior to during and now following the implementation of the Affordable Care Act (ACA). Gallup's large sample sizes have provided unparalleled ability to identify changes in the insurance status of key populations throughout the United States such as young Americans and minorities. Additionally, Gallup has also been able to analyze this data by state which is particularly important to understanding the impact of Medicaid expansion and locally managed marketplace exchanges throughout the United States as part of the ACA. In addition to tracking the insurance status of those currently living in the United States, Gallup closely monitors the percent of newly insured Americans that obtained their insurance through the exchanges, and public reaction to and support of the ACA including the perceived impact of the policy on Americans and their families. This presentation will provide researchers important insights into how the uninsured rate has changed over time, controlling for additional factors such as changes in the employment market more generally, including how these changes have occurred among important subgroups of interest such as youth and minorities.

When Numbers Aren't Enough: Supplementing Quantitative Data Collection with Qualitative Insights

Panel organizer: Jennifer Hunter Childs, *U.S. Census Bureau*

Understanding public opinion, especially as the landscape quickly changes, requires more than a standard quantitative data collection. Qualitative collections can lend a richness and depth to the data that is missing by simply looking at the numbers. On the opposite end of the spectrum, without quantitative data, a qualitative study can be unbounded and difficult to understand how the parts fit within the whole that is being studied. Combining the methods results in a study that is far richer and more complete than either one alone. This panel brings together mixed methods studies of public opinion of various topics from statistics to drones, all in the survey context. It also examines components of survey methodology, including response burden, use of administrative records for statistical purposes, and survey participation, from both qualitative and quantitative perspectives. In each of the studies, we see that what we need to know about how the public views our methodologies are best understood using a combination of qualitative and quantitative data.

What Drives Public Opinion? Drawing Causal Inference Through Open-ended Probes
Gerson Morales, *U.S. Census Bureau*

Over the years, quantitative evidence suggests that people's trust in the U.S. government has decreased (Gallup, 2014). Correlational research focusing on public opinion often shows a relationship between distrust in the government and negative press related to government events (Ceron & Memoli, 2015; Cook & Gronke, 2005). Recent events such as the Snowden/NSA scandal seem to have further influenced a decrease in trust of government data (Childs, 2014). Although correlational analyses are useful in statistics, a drawback to this type of analysis is that it does not allow us to conclude if there is a direct causal relationship between trust in Government statistics and negative government publicity. To address this gap, the U.S. Census Bureau's Center for Survey Measurement developed in-depth analyses to understand the decrease in trust in government statistics. More specifically, this research uses a mixed-methods approach to analyze how trust in government statistics has changed over time. The Census Bureau has contracted the Gallup Organization to collect nightly survey responses measuring U.S. adults' trust in the Federal Statistical System. The current contract also includes probing questions that enable us to understand the reasons for respondents' lack of trust towards government statistics. By analyzing both the qualitative and quantitative message in the Gallup Nightly Poll, we will be able to have a holistic view of respondents' perceptions. We expect that negative events disclosed by respondents in the survey will predict decreased trust in government data. Furthermore, the open-ended probe questions will provide us with root-causes that could lead to decreased trust in government statistics. Based on the results, we hope to be able to formulate communication strategies to increase trust in government data, avoid areas that could potentially trigger distrust attitudes towards government statistics, and ultimately increase response rates to Census Bureau surveys.

An E-Fax or an EHR? Evaluating the National Electronic Health Records Survey
Paul J. Scanlon, *National Center for Health Statistics*

The National Electronic Health Records Survey (NEHRS) is an annual supplement to the National Center for Health Statistic's primary ambulatory care statistical system, the National Ambulatory Medical Care Survey, and has been fielded in some form since 2008. The NEHRS is a mail survey with telephone follow-up of office-based physicians used to measure the adoption and use of electronic health records (EHR). Quantitative evaluation of the NEHRS and its metadata revealed that some questions were not performing as desired and that some questions suffered from unacceptable levels of item non-response. Specifically, questions regarding business models and health information technology (HIT) displayed higher-than-desired item missing and presumed error rates. In an effort to determine the issues behind these problems so that they could be corrected for the 2015 NEHRS, the questionnaire was qualitatively evaluated using cognitive interviewing. This presentation will discuss both the original quantitative analysis of the 2013 and 2014 NEHRS that led to the decision to complete a qualitative inquiry, the qualitative findings that helped explain some of the variability seen in the quantitative results, and the way these findings were applied to redesign the NEHRS questionnaire. Particular focus will be given to the sections of the survey that ask physicians about their practice's business practices, their patients use of EHR, and the processes used to exchange information to and from their practice to another medical practice. The analysis of the 2013 and 2014 NEHRS data revealed some differences in the validity of these domains across different types and sizes of medical practices. Cognitive interviewing revealed that physicians, particularly those at larger group or corporate practices, are not closely involved with the day-to-day business or HIT processes and therefore have difficulty responding to questions in these domains.

Using Text Analysis to Find the Meaning of Respondent Burden
Erica C. YuWright, *U.S. Bureau of Labor Statistics*

Response scales based on ordinal ratings are often used by survey designers to collect information about subjective states, such as feelings of burden from participating in a survey. Typically, these ordinal responses are summarized using quantitative analyses. But quantitative analysis alone is not always sufficient to understand such data. What if respondents do not all understand the targeted concept of the question in the same way? Text analysis of open-ended responses can reveal insights not available from quantitative data. We explore this issue within the context of respondent burden. In an online experiment (n = 545), we explored the cognitive aspects of the burden associated with survey participation. In addition to collecting ratings for quantitative analysis, we asked participants to use open-ended text fields to name activities from real life that they would rate as being Not At All Burdensome, Extremely Burdensome, and Somewhat Burdensome, which were the extreme and middle points of the same response scale they used to rate the burden of the survey they had just participated in. These open-ended responses form a corpus representing a range of conceptualizations of burden. We analyzed the text at the term level and grouped the responses to construct frequency distributions of words that allowed us to understand and systematically analyze the large amount of unstructured qualitative data. Relationships between burden themes, ratings of burden, experimental manipulations, and demographic information were analyzed. Interpretations of these themes will be presented, along with contributions to the current understanding of what defines respondent burden and recommendations on how to measure it quantitatively.

Public Opinion on Drones: A Multimodal Analysis
Jonathan Evans, *RTI International*

Small unmanned aerial systems (UAS), or drones, are quickly being developed for public service, commercial, hobbyist, and research applications as technological advancements reduce costs, expand sensor options, and simplify the flying process. Estimates vary, but perhaps half a million drones have been sold in the last 3 years, Amazon was selling 10,000 UAS monthly by December 2014, and the Consumer Electronics Association predicts a minimum global market of $1 billion for consumer drones by 2018. UAS are rapidly becoming mainstream and are no longer part of a niche market. In the face of this expanding market, our research has analyzed sentiment toward non-military drone use between 2013 and 2015. RTI has conducted 4 sentiment analysis projects in the past 3 years, including 2 surveys. Through content analysis of social media and news media, greater light is thrown on sentiment trends between the two surveys. The research has covered government, private, and academic use of UAS. Our data demonstrates unexpected trends and shifts in opinions and perspectives. The presentation will summarize these trends.

Census Knowledge and Census Participation Among Hispanics
Yazmín García Trejo, *U.S. Census Bureau*

This paper focuses on how knowledge about the U.S. Census affects Census participation among Hispanics. Research on the Census social marketing campaign in 2010 found that hard-to-reach populations, such as Hispanics, were less likely to return Census questionnaires and were also less knowledgeable about how the Census uses the data collected from individuals (Bates and Pan 2010; Evans, Datta and Yan 2014). This mirrors a larger issue concerning political knowledge, where Hispanics are one of the groups with lower levels of political and institutional knowledge in general (Delli Carpini

and Keeter 1996; Nicholson, Pantoja and Segura 2006; Pérez 2015). However, we know little about the factors associated with Hispanics' acquisition of Census knowledge and the impact of this information on non-response rates. I expand the study of the link between Census knowledge and participation among Hispanics by addressing the following questions: a) how the socioeconomic status of Hispanics influences their Census knowledge and their participation in the Census; and b) what type of knowledge may encourage Hispanics to participate in the Census. To address these research questions I analyze quantitative and qualitative evidence. First, I analyze the Census Barriers, Attitudes and Motivators Surveys (CBAMS) conducted by the Census in 2009 and 2010. Second, I explore the distribution of questions about institutional knowledge included in the Gallup daily tracking survey (2012-2013) and third, I discuss exploratory findings from cognitive interviews regarding individuals' reactions to information about the Census. Results from CBAMS suggest that the topic and level of difficulty of the Census knowledge questions influence an individual's decision to participate in the Census.

May 14, 2016
Concurrent Session J

The AAPOR Report on Address-based Sampling

Panel organizer: Rachel Harter, *RTI International*
Mansour Fahimi, *GfK Custom Research*
Cameron B. McPhee, *American Institutes for Research*
Jill Montaquila, *Westat*
Andrew Zukerberg, *National Center for Education Statistics*

Sampling from address frames that are highly dependent on the USPS files has come to be known as address-based sampling or ABS. Survey organizations can now obtain address frames or samples from address frames that nearly cover the entire population of U.S. residential mailing addresses. Vendors of ABS frames and samples typically offer value-added data or services, such as the ability to append householder name or telephone number, indicators regarding the type of address, and demographic information on occupants of the households, where available. Even though ABS evolved initially as a lower cost method of building frames of housing units for in-person surveys, its greater potential is as an alternative to Random Digit Dialing or part of a multimode mix. Any major new development in the science of surveys deserves research and documentation of current best practices. The American Association of Public Opinion Research (AAPOR) formed a task force to report on the current state of knowledge regarding ABS. The resulting report, issued in 2015, discussed technical details about the address frames and auxiliary variables, design options for using ABS, special considerations for case eligibility and the calculation of response rates and weights in the ABS context, especially in multi-mode studies. The report also discussed issues with ABS and how survey organizations deal with them, thus detailing the challenges and opportunities provided by ABS. A panel consisting of a subset of members of the ABS Task Force will summarize important issues about using ABS frames and methods, and will participate in a question and answer discussion.

Alternative Approaches to Measuring Survey Error

How Big of a Problem is Analytic Error in Secondary Analyses of Survey Data?

Guy Alain S. Aurelien, *University of Michigan*
Brady T. West, *University of Michigan*
Joseph W. Sakshaug, *University of Manchester*

Researchers from many academic fields perform secondary analyses of survey data collected from large probability samples of populations of interest. The designs of these large probability samples are often complex, rather than simple, featuring multistage stratified cluster sampling and unequal probabilities of selection. Because of the complexity of these sample designs and because many researchers haven't received the statistical training needed to analyze survey data arising from these designs, the estimates presented in research products describing analyses of these data may not fully represent the target population, ultimately yielding misleading inferences. We refer to these types of problems as analytic errors, and recent research has suggested that such errors should define an important component of the larger Total Survey Error framework. Analytic errors based on incorrect statistical methods could produce misleading descriptions of populations under study, negating federal resources dedicated to survey data collection and potentially hindering government programs and policies that rely on survey estimates. In order to examine the potential prevalence of these analytic errors in research studies presenting secondary analyses of survey data, we present a meta-analysis of 150 sampled research products analyzing survey data from the Scientists and Engineers Statistical Data System (SESTAT), which includes the Survey of Doctoral Recipients (SDR), the National Survey of College Graduates (NSCG), and the National Survey of Recent College Graduates (NSRCG). We performed detailed reviews of the methods and results described in these 150 research studies, and this presentation will describe the frequency with which apparent analytic errors occurred. Overall, the results of this study are entirely consistent with initial pilot work in this area suggesting that analytic errors may be quite frequent in these types of studies. We conclude with suggestions for survey researchers and peer reviewers alike, in addition to directions for future research in this area.

Unit Nonresponse in the 2014 Health & Retirement Study Experimental Modules

LinChiat Chang, *www.linchiat.com*
Mary Beth Ofstedal, *University of Michigan*

The Health and Retirement Study (HRS) is a longitudinal panel survey representing Americans over age 50. After completing the 2014 HRS core interview, around 18,000 respondents were randomly assigned to 1 of 11 modules of supplemental questions, yielding approximately 1,600 respondents per module. However, not all HRS respondents who were assigned to a module completed it. Nonresponse increases the risk of bias in survey estimates. The risk of systematic bias is cause for concern when core substantive measures are associated with nonresponse propensity. Conversely, nonresponse may be ignorable if it is random and not associated with core substantive measures. This analysis will address four main questions: 1. How representative are module subsamples? Module subsamples differ slightly from population demographic parameters, in part by design because modules are administered only to non-proxy respondents who can answer for themselves. In addition, respondents who were willing to complete the module questions consistently scored higher on physical, cognitive, and emotional functioning than those who refused to complete the modules. 2. Did unit nonresponse vary by module topic? Nonresponse rate did not vary significantly across the 11 modules; and the completed samples exhibited similar demographic profiles, contactability and cooperativeness. 3. Is nonresponse propensity

systematically associated with substantive variables? Using paradata on respondents' cooperativeness and contactability to simulate nonresponse propensity, we assess the extent to which nonresponse propensity is associated with substantive module variables. We demonstrate how the impact of nonresponse may be ignorable on certain module variables; whereas other module variables are more susceptible to nonresponse bias. Bonferroni correction is applied. 4. Does nonresponse propensity change relationships between substantive variables? Multivariate models across 3 module topics of arts participation, healthcare technology, and alcohol intake will be fitted to determine whether nonresponse propensity moderates the strength and direction of established relationships between substantive variables.

Using the Planning Database to Plan and Improve the Survey Process
Kathleen M. Kephart, *U.S. Census Bureau*

The U.S. Census Bureau's Planning Database (PDB) is an easy to use, publicly available data set containing current American Community Survey (ACS) 5-year estimates and 2010 Census data. It is available at both the block group and tract levels of geography. In addition to demographic estimates and decennial operational data, the PDB also contains the Low Response Score, a model based estimate that assigns a score for survey response propensity. The availability of the PDB at such a low level of geography provides an opportunity to improve several aspects of the survey process, from data collection planning to analysis. Since the creation of the PDB, prior to the 2000 Census, the Census Bureau has used the database to plan for field hiring efforts and to target areas with historically low response rates for the decennial census. As it is publicly available, the PDB is also a potentially valuable tool for non-governmental survey practitioners. The PDB can be used to locate tracts or block groups with desired characteristics (e.g. seniors, children, Hispanics, languages spoken, poverty rates, health insurance coverage rates, etc.), to inform sample design, and the allocation of resources. During the analysis phase it can further inform findings about a target population, create GIS maps, enhance reports, or be used to build models. For the instances when a data set or sampling frame does not match easily to a Census tract or block group level, but instead has only a Zip Code geography, we will demonstrate how to use publicly available data to create a Zip Code Tabulation level PDB. We will then compare the distribution of aggregated response propensity scores across geographies. We will also present several relevant examples for survey practitioners using the current PDB for the survey planning and analysis process.

Latent Class Assessment of Measurement Error in a Large Scale Panel Survey
Brian Meekins, *U.S. Bureau of Labor Statistics*
Clyde Tucker, *American Institutes of Research*

While most survey information is measured with error, efforts to detect the amount of measurement error in survey estimates are somewhat limited and tend to focus on external benchmarks or comparisons with other surveys. Unfortunately, it is frequently the case that there are no benchmarks for survey estimates. In this case, we may rely on a variety of paradata or demographic information to roughly assess the amount of measurement error. These indicators are best treated as indicators of an underlying or latent variable that assesses the potential measurement error in the particular estimates or the interview as a whole. Building on previous research (Tucker et al. 2011) this work develops such a measurement error latent construct using data from the Consumer Expenditure Interview Survey – a large scale, rotating panel survey conducted by the Bureau of Labor Statistics. A large set of indicators are used that incorporate administrative data such as mode of administration, use of interview aids (including bills, receipts, and other records), and contact history, as well as interviewer assessments of

cooperation and quality, and other more commonly used indicators (such as item missingness, length of interview, patterns of reporting). We then examine the demographic profile of each class of measurement error, as well as the consistency of the latent variable over the course of the panel.

Making Twitter Glitter in Survey Research

Finding Wheat in the Chaff: Using Twitter to Inform Social Science Research

Ian S. Thomas, *RTI International*
Scott Novak, *RTI International*
Gary Zarkin, *RTI International*
Martin Duparc, *RTI International*

Social media has emerged as an important tool in the social scientist's arsenal in capturing information about knowledge, attitudes, and behaviors. Because the data are sourced directly from the participant in near real-time, this method holds tremendous opportunity as an early warning system that can allow researchers to detect early signals in population dynamics. Social media tracking has been used to track a diverse range of issues in public policy and public health. Unlike survey data which standardize and validate survey questions so as to reduce interpretative bias, social media data are obtained in an unstructured manner directly from the participant. Therefore, novel methods are needed to rapidly organize textual data in a way that can preserve the message's meaning and context. This study presents our work in developing a tweet tracker for medical marijuana in the United States. Because of the sensitive and perhaps illicit nature of marijuana use, twitter users often use code to hide meaning from law enforcement or authorities. In order to investigate this further RTI developed tools for collecting, classifying, and analyzing tweets. Because of the ever-changing nature of Twitter, a process was developed for utilizing both automated discovery and specialist feedback for refining the set of tweets collected on a daily basis. The classification process was developed to take advantage of subject matter experts for creating a classified set of tweets used to train a classification model. The analysis tools were created to allow researchers to more efficiently extract meaning and content from a large set of tweets. This paper will detail the processes and tools that were used to distill data for tracking new themes related to marijuana, including motivations for use. This talk will conclude with a series of recommendations, best practices, and challenges in collecting, classifying and analyzing Twitter data.

Tracking Candidate Favorability on Social Media: Comparing Twitter Data with a Rolling Cross-section

Jake R. Dailey, *University of Michigan*
Josh Pasek, *University of Michigan*

A number of recent studies have attempted to track elections using Twitter data. Despite early successes, a recent meta-analysis argues that attempts to do so perform no better than chance. Specifically, Twitter data and survey data appear to yield unrelated conclusions about election dynamics. We propose that failures in this arena may be due to differences between Americans at large and the population of Twitter users. Hence, when trends between these two sources diverge, it might say less about the quality of Twitter data and more about the population using the site.
The current study examines whether the conclusions derived from Twitter are more similar to those of surveys if the survey data is restricted to the kinds of people who use Twitter. To accomplish this, we combine three sources of data: the sentiment of Tweets about Obama and McCain in the lead-up to the 2008 U.S. Presidential election, favorability of those same candidates as derived from the 2008 National

Annenberg Election Survey (a rolling cross-sectional study), and a post-election Pew dataset that included questions about Twitter use. The post-election Pew dataset was used to derive a set of weights estimating the probability that each individual would be a Twitter user. These weights were then applied to the Annenberg data to generate estimates of candidate favorability among the kinds of people who used Twitter in 2008.

In contrast to expectations, over-time correlations between survey self-reports from the kinds of people who used Twitter and Twitter sentiment were not substantively higher than between all survey respondents and Twitter sentiment. This implied that the principal reason that Twitter data did not mirror survey data was not a function of differences between Twitter users and survey respondents but instead represented a fundamental difference in the nature of the data derived from each source.

Addressing the Demographic Bias on Twitter

Soroush Vosoughi, *MIT Media Lab*
Mina Soltangheis, *MIT Media Lab*
Deb Roy, *MIT Media Lab*
Russell Stevens, *MIT Media Lab*

In recent years, the ubiquity, ease-of-use and accessibility of social media have made it one of the main sources of data about human behavior. Specifically, Twitter, given its public nature, has become one of the main sources of data used to study human behavioral and social dynamics. But as is the case with any information channel, there are inherent demographic biases with Twitter. Given the thousands of scientific articles using Twitter data, one would expect this bias problem to have already been addressed. This is surprisingly not the case. In this paper, we propose a method for correcting the distorted demographics on Twitter. We start by automatically detecting the position of a set of Twitter users towards several political issues such as abortion, gun rights, etc. This is done through a state-of-the-art "stance" system that we have developed. We then look for the distribution of these political positions in polls conducted in the real world - specifically, distribution across different demographics (e.g., what percent of males aged 18-25 believe in made global warming). This process leaves us with three pieces of information - the distribution of political positions on Twitter, the distribution of these same positions in the real-world, and the demographic breakdown of the people polled in the real-world. We can then use a Bayesian approach to infer the demographic of the users on Twitter. Additionally, any other contextual data that is made available to us through Twitter, such as the geo-location of their users, can be used to set the priors for the Bayesian model, thus improving its performance. The model can be evaluated by using Twitter's polling capabilities, that has been made available to the public in recent months. Through this tool, we can directly collect demographic information from a few hundred Twitter users for evaluation.

Information Exposure and Sharing Behavior of e-Cigarette Users: Do Survey Responses Correlate with Actual Tweeting Behavior?

Annice Kim, *RTI International*
Jamie Guillory, *RTI International*
Brian Bradfield, *RTI International*
Paul Ruddle, *RTI International*
Y. Patrick Hsieh, *RTI International*
Joe Murphy, *RTI International*

E-cigarette use rates among adults have increased substantially, from 1.8% (2010) to 13% (2015). This increase in popularity is driven partially by marketing and communication about e-cigarettes on Twitter. Twitter provides opportunities for exploring how e-cigarette users are exposed to and share information about e-cigarettes. This, in turn provides insights into marketing and use of e-cigarettes and potentially effective prevention communications. We compared self-report data from surveys to Twitter data to validate whether e-cigarette users were truly exposed to and shared information about e-cigarettes on Twitter. Participants were recruited via Twitter ads (N= 265) and online panel (N= 116), provided Twitter handles, and answered questions about information sharing and exposure on Twitter, e-cigarette use and demographics. We used keywords to gather tweets about e-cigarettes from 1 year preceding survey data collection, both from participants and the people they follow. The sample was predominantly white (63%) and male (59%) (Mean age= 32, SD= 12). One third of participants had a Bachelor's degree or more. On average, participants were exposed to 4,962 e-cigarette Tweets (Median= 165, Range= 0-558,206) from 78 friends (Median= 37, Range= 0-1,013), and shared 23 e-cigarette Tweets (Median= 0, Range= 0-4,651). Self-reported exposure to e-cigarette information on Twitter was closely tied to observed behavior on Twitter, such that 97% of those who reported seeing e-cigarette information on Twitter in the past 3 months had friends who tweeted about e-cigarettes. In contrast, only 38% of e-cigarette users who reported sharing e-cigarette information in the past 3 months on Twitter tweeted at least once about e-cigarettes. Findings indicate that researchers should be careful about relying solely on self-reported responses for measuring information sharing behavior on social media. Demographics and e-cigarette use behaviors of those who share and are exposed to a higher number Tweets about e-cigarettes will be discussed.

Regularized Correspondence Analysis: An Ideal Point Model for Massive, Sparse Data with an Application to Twitter

Solomon Messing, *Pew Research Center*

We construct a scalable, dynamic measure of ideology for citizens on social media using preferences revealed in the course of following Twitter accounts. We assess the measure's convergent and predictive validity against the accounts of well-known politicians and common strings of text appearing in the content. We perform a regularized correspondence analysis on the incidence matrix of users and those they follow. This method scales to the large data sets necessary to produce measures with broad coverage, where more complex methods such as item-response theory (IRT) models fail. We use the method to examine polarization online in the context of the 2016 campaign.

Changes and Constants Over Time

Continuing to Explore the Relation of Economic and Political Conditions with Government Survey Refusal Rates, 1960 to 2015

Luke J. Larsen, *U.S. Census Bureau*
Joanna Fane Lineback, *U.S. Census Bureau*
Benjamin M. Reist, *U.S. Census Bureau*

In this presentation, we examine environmental and political influences on Current Population Survey (CPS) refusal rates. Government statistical agencies are operating in a difficult survey climate. Response rates for a number of major government surveys have experienced rapid declines. Much of the research into causes of declining response rates has focused on micro-level influences (e.g., interviewer workloads). We are interested in examining possible macro-level influences. As such, a reasonable starting point is extending the work by Harris-Kojetin and Tucker (1999). The authors used CPS data and a time-series regression approach to examine environmental and political influences (unemployment rates, presidential approval ratings, inflation rates, and consumer sentiment scores) on CPS refusal rates over the period 1960-1988. Now over some 25 years later, we are interested if the authors' findings still hold over the period 1960-2015.

Survival Analysis: Estimating Attrition Rates in the NLSY97

Isabella Velasquez, *NORC at the University of Chicago*
Holly Hagerty, *NORC at the University of Chicago*

Survival analysis explores the effect of time-varying and invariant covariates on the length of time before a well-defined event, such as extinction or death. Previous research using survival models in conjunction with longitudinal datasets has focused on measuring the duration until respondents reach certain outcomes—graduation or dropout from high school, changing employment status, or transitioning from licit to illicit drug use (Radcliffe 2014, Malone 2015, Wu 2011). In contrast, the objective of this paper is to model attrition risk over time for respondents in the National Longitudinal Survey of Youth, 1997 (NLSY97). We are interested in predicting the amount of time until the first year an NLSY97 respondent attrites, or is not interviewed for a round. We fit several survival models to the NLSY97 data for years 1997 through 2011, representing 14 rounds of annual data collection. We compare survival model estimates to estimates obtained from conducting a logistic regression of attrition on a set of observable baseline characteristics, including gender, race, baseline household income, and parental education levels. We test the accuracy of our predictions by comparing estimates of Round 15 attrition derived from the survival and logistic models with true Round 15 attrition data. Although our relatively short time range weakens the predictive power of our model, survival analysis provides useful insights on dropout rates based on a person's characteristics. In particular, we find higher probabilities of attrition for male respondents, Hispanic respondents, and respondents with lower average levels of parental education. These findings can help longitudinal survey researchers identify which subgroups of respondents might be at a particular risk of attrition based on observable characteristics, as well as the rate at which respondents might refuse to participate. We suggest that targeted retention strategies can improve the retention of respondents who fall into "high attrition risk" subgroups.

Joining & Staying In: Motivations for Panel Participation

Jenna Levy, *Nielsen*
Erin Wittkowski, *Nielsen*
Adam Gluck, *Nielsen*
Arianne Buckley, *Nielsen*
Courtney Mooney, *Nielsen*

Recruiting and retaining participants in a long-term panel study poses unique challenges, requiring the development of recruitment methods that appeal to respondents' motivations and strategies to maintain long-term engagement. This research focuses on improving cooperation and retention rates for long-term panels. There have been many studies about why respondents may choose to participate. According to leverage-salience theory (Groves et al, 2000), respondents may be more inclined to participate in a survey if they are made aware of attributes that interest when they are first introduced to the survey (e.g. topic or incentives). Additionally, social exchange theory as applied to surveys posits that people will choose to participate in a survey or not based on their individual cost-benefit analysis (Blau 1964; Homans 1958). However, the environment has radically changed since these theories were first introduced. This research seeks to understand if these theories still hold true, and if there are different factors affecting participation today. We conducted a series of mixed-method studies to explore the driving factors behind joining and participating in a panel. These studies focused on what influences one's decision to join a panel, as well as what motivates panelists' ongoing cooperation with a daily task. Furthermore, this research examines both the intangible (e.g., representing one's community) and tangible reasons (e.g., monetary incentives) for panel participation. Methods include web and telephone surveys and focus groups among former Nielsen panelists. Analyses of these studies will look for patterns with regards to intrinsic and extrinsic motivations for panel participation by demographic characteristics to ultimately inform recruitment and retention strategies. Analyses will explore potential generational differences, as well comparisons between people who participate and those who do not. Also, analyses of Nielsen's studies among former panelists will examine influences on participation by sample performance, panel tenure, and reasons for leaving a panel.

Factors Influencing Recurrent Consent Requests for Biomeasures in a Longitudinal Survey of Older Adults

Brian M. Wells, *University of Michigan*

Longitudinal studies that collect physical and biological measures (biomeasures) have not been aided by past research focused only on a single wave of data collection where a consent request was made. As these longitudinal studies continue making repeated consent requests for biomeasures, it is vital to understand how a respondent's previous consent (or non-consent), changing health, and the survey design impacts the probability of obtaining future consent. This study looks at consent to collect dried blood assays and various physical measurements across four waves of the Health and Retirement Study (HRS) from 2006 to 2012. Panelists were asked to provide various biomeasures every four years (every other wave) during face-to-face interviews. Models examined in this paper assess the impact of socio-demographic variables, recent changes in health, interviewer attributes (including continuity of interviewer across waves), and previous interview resistance as influential predictors of future waves of consent for biomeasures. Of particular interest are factors that identify panelists who deviate from their previous consent status: those who consented in a preliminary wave but not in a later wave and those who refused to consent initially but consented in a later wave. This paper identifies factors associated with repeated biomeasure consent beyond those generally associated with panel attrition, consider how

such factors can influence and improve interviewer training, and discuss the impact such findings will have on post-survey adjustments and adaptive survey design.

Changes in Labor Market Behavior Due to Panel Conditioning in a German Panel Study

Ruben L. Bach, *Institute for Employment Research (IAB)*
Stephanie Eckman, *RTI International*

Panel conditioning refers to changes over time caused by participation in a panel survey. Changes may occur in respondents' behavior (for example, participants in an election panel may be more likely to vote) and/or in their reporting of behavior (for example, respondents become more trusting of interviewers and report more socially undesirable behavior). Changes in behavior, however, are usually much harder to identify; some researchers have even denied that changes in behavior are possible. Using administrative data linked to a large panel survey, we address the challenges posed to previous researchers and analyze changes in respondents' behavior due to repeated participation in the panel, specifically whether responding in a labor market survey reduces the amount of time needed to find a job. In previous research using the same linked data set, we used propensity score weighting models to show that participation in three waves of the survey leads to increases in the take-up of federal labor market programs. We build on this research by extending our work to a new variable: job search time, which is of much greater substantive interest to the international community. Preliminary results show that panel respondents find a job faster than they would have had they not participated. Moreover, we analyze whether these panel conditioning effects vary with the number of waves. That is, do we see a small effect after one wave, and increasing effects after each additional wave? We interpret these results as evidence that surveys can induce changes in behavior among respondents. These results will be of interest to all researchers who collect or analyze panel survey data.

The Mechanics of Election Polls

RAND 2016 Presidential Poll Baseline Data

Michael S. Pollard, *RAND Corporation*
Joshua Mendelsohn, *RAND Corporation*
Alerk Amin, *RAND Corporation*

This presentation will introduce the 2016 Presidential Poll baseline data in the RAND American Life Panel (ALP). The ALP consists of a panel of about 6,000 U.S. respondents ages 18 and older who regularly take surveys over the Internet. An advantage over most other Internet panels is that the respondents to the ALP need not have Internet access when they are initially recruited and thus the panel can be based on a probability sample of the U.S. population. The ALP will collect a comprehensive baseline survey in December, including questions on party affiliation, primary voter choice, probabilistic polling, several topical issues (e.g., immigration, Iran, the economy), head-to-head matches (Clinton vs. Trump, Clinton vs. Bush, Sanders vs. Trump, etc.), and candidate traits that matter. Respondents will be re-interviewed throughout the election season. Responses will also be linked back to the respondents' original 2014 midterm, 2012 and 2008 presidential election data where possible. The baseline data will be collected well in advance of the major party debates, caucuses, and conventions, facilitating clear tracking of respondent attitudes over the election cycle. Results of the baseline, and potentially of subsequent surveys collected by the time of the presentation, will be presented as a way of stimulating discussion.

All data collected are available for free to researchers. The ALP is also a service for researchers to field their own questionnaires and experiments.

Cross-national Comparisons of Polling Accuracy
Jacob Sohlberg, *University of Gothenburg*
Mikael Gilljam, *University of Gothenburg*

The recent failures of pre-election polls in countries such as the UK, Poland and Israel illustrate how difficult polling can be. While such errors receive much publicity, not all countries are experiencing the same problems with accuracy. The purpose of the paper is, first, to conduct cross-national comparisons of polling accuracy and, second, to better understand the underlying causes of variability in polling accuracy. To make polling accuracy, our dependent variable, comparable across countries, we are focusing on polling estimates for the two largest parties and including all smaller parties in a category for other. For each country, we are combining multiple polls, conducted close to the election, to form average polling estimates. The mean level of support according to the polls is then compared with the election results to produce a combined metric of polling accuracy. Our main independent variables are trust in other people, level of turnout and vote share of anti-establishment parties. We predict that higher polling accuracy is influenced by (1) a higher level of trust in other people because it makes individuals more likely to participate in surveys and give honest answers, (2) a higher level of turnout because pollsters are then less reliant on challenging likely voter models and (3) a smaller vote share for anti-establishment parties because voting for such parties is associated with more voter instability and potentially more social desirability bias. In our multivariate analysis, we control for factors such as sample size and probability-based versus opt-in sampling.

Rational Giving? Measuring the Effect of Public Opinion Polls on Campaign Contributions
Dan Cassino, *Fairleigh Dickinson University*

During American presidential elections, small contributions ($250 or less) now make up nearly half of all of the money raised by candidates during primary and general elections. But are these small donors giving money in a rational way? This paper uses the universe of campaign contribution data available from the FEC in conjunction with daily polling averages and content analysis of news coverage of the candidates in order to isolate the effects of changes in the candidate's standing in the polls on the propensity of donors (grouped by the amount of money donated) to contribute money towards that candidate. During the 2012 elections, the results show that small contributors pay attention to the candidate's standing in the polls, but not the trajectory of those poll numbers, while larger donors put a greater weight on whether the candidate is increasing or decreasing in the polls, making them far less likely to contribute to a candidate on their way down. So, while public opinion data matters to everyone, larger contributors are much more likely to use polling data rationally when making a contribution decision, which is perhaps one reason for their perceived outsize influence on the electoral process. Finally, the paper examines early data from the 2016 primary elections to see if the same patterns still hold in the current election.

Health & Health Care: Measurement, Methods

It's Getting Late: Improving Completion Rates in Hard-to-Reach Samples

Jennifer Cantrell, *Evaluation Science & Research, Truth Initiative*
Randall K. Thomas, *GfK Custom Research*
Alexandria A. Smith, *Evaluation Science & Research, Truth Initiative*
Valerie F. Williams, *Evaluation Science & Research, Truth Initiative*
Elizabeth C. Hair, *Evaluation Science & Research, Truth Initiative*
Donna Vallone, *Evaluation Science & Research, Truth Initiative*

Increasing response rates during late-stage fielding of large surveys is a common challenge in survey research. This study presents 2 interventions to improve late-stage survey response in a probability-based sample of 12000 15-21 year old participants recruited for an online panel to assess smoking behaviors and attitudes for 6 survey waves.   To boost completion rates in our Wave 2 study (Winter, 2015), we implemented an experiment (n=4467) to test 2 primary manipulations–increase minimum incentive from $20 to $30, $40, or $50 and randomly assign to an enhancement (control condition; sweepstakes; IVR calls; phone prompts; postcard reminder; combo of all). The completion rate average was 59.2%. The increase in total complete rate due to the experiment was 6.2%, with 699 of the 4467 participants assigned to the experiment completing Wave 2. The primary improvement occurred due to the postcard group with secondary influence by sweepstakes and phone prompt. Our second study reports on an intervention to target non-responders in Wave 3. We had 7761 completes out of a sample of 13014 sample (ABS+phone recruits) before implementing a final effort to obtain additional completes–enclosing a $5 non-contingent incentive with a letter in an envelope sent by postal mail to the non-responders (n=5253). We had 8139 completes before the intervention (Sept. 18) and 8795 completes 29 days later, of which we attributed approximately 520 completes due to the manipulation (versus an estimate of 106 completes without any reward intervention). Without the intervention, we estimated a completion rate of 63.3%; with the intervention, we had a completion rate of 67.6%. The total cost for each additional complete represented about $85 more to obtain. We discuss the results in terms of the costs and benefits of late term interventions on completion rates.

Measuring Trust in Medical Researchers: A Comparison of Agree-disagree and Construct-specific Scaling Formats

Ian F. Wall, *University of Wisconsin-Madison*
Jennifer Dykema, *University of Wisconsin - Madison*
Dorothy Farrar Edwards, *University of Wisconsin-Madison*

Survey scales that measure subjective constructs historically rely on agree-disagree (AD) items. However, several sources of bias, including extreme response bias and acquiescence bias, pose a significant threat to the validity and reliability of this method of measurement. Several methods for reducing these biases have been proposed and tested, for example item reversals, anchoring vignettes, and construct-specific (CS) response categories. Scholars have argued that CS items clarify underlying dimensions that AD questions leave implicit, and several studies have shown that responses to CS items generally provide higher measures of reliability and validity (Saris, Revilla, Krosnick, & Shaeffer, 2010). Yet there is relatively little research that compares AD and CS items during the response process, so that we may better understand specific sources of error and develop targeted solutions. We address this by using cognitive interviewing techniques (Willis, 2004) to provide insight into how participants process AD and CS items, how they interpret the intent of the question, comprehend terms, retrieve information

from memory, and map their answers to response categories. We provide preliminary results from 64 cognitive interviews with a quota sample of African Americans, Native Americans, Caucasians, and Latinos. Interviews were conducted to test questions about barriers and facilitators to participation in biomedical research. Participants were randomized to either an AD or CS version of an 11-item scale measuring trust in medical researchers. We compare response latencies, indicators of response processing difficulties (e.g., requests for clarification), and indicators for reliability and concurrent validity for each question type. In addition, answers to questions and probes were analyzed qualitatively. Overall, we find that CS questions do not necessarily improve all aspects of the response process. Based on our experience with this small study, we offer several suggestions for developing and testing CS questions.

Health and Health Behavior Differences in Four Communities

John M. Boyle, *ICF International*
Eric Miller, *ICF International*
Lew Berman, *ICF International*
Jamie Dayton, *ICF International*
Ronaldo Iachan, *ICF International*
Melanie Courtright, *Research Now*
Kartik Pashupati, *Research Now*

Most public health programs are implemented and evaluated at the community level, but few public health surveillance systems generate community measures. National health surveys lack the sample size to provide estimates at the community level. Conducting parallel community-level health surveys could substantially augment the benefits of national surveys, and operationalize findings on the local level. For example, community level surveys could help identify areas to target resources for cancer survivors or programs to increase cancer screening. However, costs for community-level probability surveys have been a barrier to adoption of community health surveys. In previous research, we have demonstrated that community-level demographic characteristics and key health indicators from web panel surveys are comparable to measures from national surveys. The objective of this analysis is to assess the ability of web panel surveys to measure the burden of cancer and utilization of cancer screening at the community level. In this analysis, we will estimate self-reported cancer diagnoses and utilization of cancer screening tests (for breast, colorectal, prostate and cervical cancer) from web panel surveys and make comparisons between four communities and to national data. We use population web panels drawn as Census balanced samples to conduct community health surveys among 500 or more adults each in New York City, Cleveland, Seattle and Los Angeles. The community findings will be compared to the national Health Information National Trends Survey (HINTS). We believe that the differences in health and health related behaviors between communities in this study offers valuable insights into needs, opportunities and outcomes of great importance to national and local health care programs.

The Effect of Individual vs. Group Response on Instrument Reliability for a Medical Practice Survey

Nancy Duda, *Mathematica Policy Research*
Dmitriy Poznyak, *Mathematica Policy Research*

In 2012, the Centers for Medicare & Medicaid Services launched the four-year Comprehensive Primary Care (CPC) Initiative, designed to improve patient outcomes by improving primary care delivery in seven regions across the United States among nearly 500 practices. As the healthcare field continues to test

different patient care models, including CPC, it is important to measure practice transformation progress and to identify specific areas for improvement. Mathematica developed the CPC Initiative Practice Survey, which included 26 questions adapted from the existing PCMH-A instrument and 15 items designed by Mathematica, to measure practices' approaches to delivering primary care. Instructions recommend that the practice manager complete the survey with input from others in the practice. The survey collects information on who completed the questionnaire, whether and how group input was gathered to complete it, and which staff provided input. Mathematica has administered the survey three times since fall 2012 to practice managers of CPC practices. For each round, we look at the responses from the CPC practices and examine changes in the reliability of survey responses depending on whether the responses drew on input from only the practice administrator, input from additional staff including clinicians, or input from additional staff excluding clinicians. We first estimate a set of models to measure reliability using multi-group Confirmatory Factor Analysis. We use Item Response Theory analysis to further examine changes in reliability performance of individual items and scales, depending on the type of survey response input. Our hypothesis is that group staff input, particularly inclusion of clinician input, results in improved reliability. This would suggest that establishment surveys—including primary care practice surveys—should be completed in collaboration with the staff to maximize the reliability of survey measures.

Contrasting Measures of Health Insurance Literacy and Their Relationship to Health Care Access

Kathleen Thiede Call, *State Health Access Data Assistance Center*
Giovann Alarcon, *State Health Access Data Assistance Center*
Alisha Baines Simon, *Minnesota Department of Health*
Sarah L. Hagge, *Minnesota Department of Health*
Karen A. Turner, *State Health Access Data Assistance Center*
Stefan Gildemeister, *Minnesota Department of Health*

The Affordable Care Act resulted in significant increases in health insurance coverage among the U.S. population. However, when coverage does not translate into access to care, expected gains in population health and health equity are not achieved. Low health insurance literacy (HIL) is a likely driver of underuse of insurance for preventive care. HIL is defined as the extent to which people have access to insurance information and the ability to process and use that information. Using the 2015 Minnesota Health Access Survey (a dual frame telephone survey) we compare patterns of results for two distinct measures of HIL: one focusing on understanding insurance terminology and a second that assesses proactive use of coverage. Participants with public and private insurance were randomly assigned to one of these measures. Compared to people with public insurance those with private insurance demonstrate greater confidence understanding of financial terms (e.g., deductible, co-payments), provider network and open enrollment, as well as the four items in the proactive use scale (e.g., is a doctor is in-network or what the plan will and will not cover before seeking services) and the aggregate score. Regardless of insurance type, based on understanding terminology those with low HIL are more likely to forgo care due to cost and report less confidence in getting needed health care, however, the opposite is true for proactive use of insurance and forgone care. This suggests these measures are tapping different concepts. We explore associations between HIL and access in greater depth accounting for demographic, health status and utilization in multivariate analysis. Early results indicate that HIL is higher for privately insured using either measure, yet for both insurance types low HIL diminishes confidence in getting needed care using either measure and the results are mixed for forgone care depending on the HIL measure used.

Interviewer Effects

Interviewer Effects on Response Latencies in a Face-to-Face Interview Survey

Patrick Sturgis, *University of Southampton*
Gabrielle Durrant, *University of Southampton*
Olga Maslovskaya, *University of Southampton*
Ian Brunton-Smith, *University of Surrey*

It is well known that interviewers are key in determining the quality and cost of data collection in household interview surveys. They influence the likelihood that sample members will respond and, therefore, the bias and precision of estimates. They also affect the answers that respondents give in ways that can make them more (or less) accurate than they would have been in the absence of an interviewer. These effects have been robustly demonstrated in the existing literature over many years. In this paper we consider a different and less well studied outcome; the extent to which interviewers affect the time respondents take to answer questions. Because interview length is used as a key indicator of both response quality and fieldwork costs, the answer to this question is consequential. We apply a cross-classified mixed-effects location scale model to response latencies in the UK Household Longitudinal Study. The model extends the standard two-way cross-classified random-intercept model by specifying the residual variance to be a function of covariates and additional interviewer and area random effects. This extension enables us to study interviewers' effects not just on the 'location' (mean) of response latencies, but also on their 'scale' (variability). We can therefore address questions of how variable response times are across interviewers, as well as by how much interviewers push the average time take to answer questions up or down. By linking the response latency data to an independent survey of interviewers, we are also able to model these interviewer effects as a function of their demographic, work experience, attitudinal, and personality characteristics.

Interviewer Influence on Interviewer-Respondent Interaction During Battery Questions

Beth Cochran, *University of Nebraska-Lincoln*
Kristen Olson, *University of Nebraska-Lincoln*
Jolene D. Smyth, *University of Nebraska-Lincoln*

In interviewer-administered surveys, the presentation of battery questions is often not standardized across the interviewers. Due to the design of battery questions, where a list of items relates to the same question stem and response options, interviewers are often allowed the discretion of when to repeat the stem and/or response options (Dillman et al., 2014). By allowing this discretion, the interaction between interviewers and respondents may vary across items within a battery. The data for this study is from the U.S./Japan Newspaper Opinion Poll, a telephone survey of U.S. adults conducted by Gallup during November 2013 (AAPOR RR1=7.4%). This survey contains four batteries with the number of items ranging from six to fourteen. We use behavior codes to examine the likelihood of interviewer assistance and respondent requests for assistance. Each respondent answered all of the items, allowing us to use a multilevel logistic regression model with three levels to take into account the between respondent and between interviewer variance. We included two item-level predictors, the item content and order of presentation, two respondent-level predictors, age and education, and two interviewer-level predictors, age and tenure of employment. Preliminary analyses indicate that the behaviors of interviewer assistance and respondent requests were more likely to occur during items presented at the beginning

of the battery. We will conclude with suggestions for designing battery questions and training interviewers regarding the implementation of such questions.

Do Interviewer-Respondent Interactions in CAPI and CATI Interviews Show a Difference in Rapport?

Yfke P. Ongena, *University of Groningen*
Marieke Haan, *University of Utrecht*

Various studies have shown that social desirability bias and satisficing are more prevalent in CATI than in CAPI surveys. Although this difference has theoretically been explained in terms of rapport (Holbrook et al 2003), it has not systematically been studied whether interviewer-respondent interactions in CATI and CAPI surveys indeed show a difference in rapport. In our study we aim to fill this gap. We analyzed 60 CATI and 54 CAPI interviews that originated from a mixed-mode experiment using the European Social Survey questionnaire (Haan 2015). Analysis was based on a coding scheme developed by Garbarski, Dykema and Schaeffer (forthcoming, 2016), who define rapport in terms of responsiveness by interviewers and engagement by respondents. We found mixed differences with respect to behaviors related to rapport. For example, interviewer laughter appeared to be more common in CATI than in CAPI, but apologetic utterances such as 'sorry' occurred equally often in both modes. Furthermore, a significant difference was found in the number of words uttered. Question-answer sequences contained on average two more words in CATI than in CAPI. This effect is partly explained by the fact that for many questions in the CAPI survey show cards were used, and in those cases extension of interaction in CATI interviews is due to less efficient communication about response alternatives. Further analysis of extended interactions showed that respondents in CATI had more difficulty in formulating their response and had more difficulties with question wording than in CAPI. These task-related issues may contribute to the effect of decreased trust and motivation of respondents in CATI interviews, and may subsequently explain the increased level of satisficing and social desirability bias in this survey mode compared to CAPI.

Decomposing the Interviewer Variance Introduced by Standardized and Conversational Interviewing

Brady T. West, *University of Michigan*
Frederick G. Conrad, *University of Michigan*
Frauke Kreuter, *University of Maryland, JPSM*
Felicitas Mittereder, *University of Michigan*

Recent research in survey methodology has attempted to decompose the interviewer variance introduced in interviewer-administered surveys into its potential sources, using the Total Survey Error (TSE) framework. These sources include sampling error variance, or a breakdown of interpenetrated assignment of sampled cases to interviewers; nonresponse error variance, or variance among interviewers in the types of cases recruited; and measurement error variance, or variance among interviewers in mean response errors. No research to date has attempted to compare this decomposition for different interviewing techniques. In this presentation, we will compare the interviewer variance introduced by conversational (CI) and standardized (SI) interviewing. We will present results from an experimental study mounted in Germany, where a national sample of employed individuals was measured on a variety of cognitively challenging items related to employment history by interviewers randomly assigned to receive training in the use of either CI or SI. This study featured an area sample and an interpenetrated sample design, enabling estimation of interviewer variance, and the

presence of high-quality administrative information on the sampling frame also enabled estimation of interviewer variance in the various sources of error. Employing analytic techniques designed to compare the various components of interviewer variance between the SI and CI groups, we will demonstrate that the interviewer variance introduced by the two techniques for selected survey items does appear to be driven by interviewer variance in different error sources. We will conclude with suggestions for practice based on the study results and directions for future research in this area.

Maintaining Data Quality During Fieldwork in International CAPI Surveys

Panel organizer: Martin Wulfe, *MWulfe Consulting*

This panel addresses the special issues related to CAPI fieldwork in the developing world including sampling and listing challenges, case management and fieldwork supervision, data collection hardware and software, data transmission and synchronization, and planning and logistics. Presentations will include cases in specific surveys and comparative studies across countries and cultures, including the use of paradata to maintain data quality.

Item Response Time in Household Surveys in Developing Countries: A Multinational, Multiregional and Multicultural Perspective
Michael Wild, *World Bank Group*
Michael Lokshin, *World Bank Group*

In the recent years survey paradata has received increasing attention in research on survey methodology, as an important element in the quality management of survey data as well as in respect to total survey error. One important element in survey paradata are the response times at the unit as well as the item level. So far research on response times and its determinants has focused almost exclusively on paradata generated by CATI or web surveys in developed countries, due to the limited availability of this kind of data in other survey modes or in developing countries. Through an increase in CAPI data collection in general, and in developing countries in particular through the World Bank's Survey Solution CAPI package, this data is now also available for developing countries, where survey data and the related quality is of particular importance, as these are often the only source to produce estimates for the population at large. As a lot of estimates derived from survey data for developing countries are benchmarked against each other or act as indicators for the MDGs or SDGs, comparability of survey data across countries is a crucial element. Besides other factors influencing comparability in 3M surveys data quality is one of them. For this purpose we analyze paradata for three countries – Zambia, Benin, and Bhutan respectively - and derive implications for survey planning and management from it. We also argue that standardization and harmonization of paradata across countries could be an important element to improve survey data collection, the resulting data quality and the comparability of the derived estimates across developing countries.

Report on Using CAPI by the MLFD in Tanzania
Niwael Mtui, *Ministry of Livestock and Fisheries Development, Tanzania*
Michael Rahija, *Global Strategy to Improve Agriculture and Rural Statistics (FAO)*

Decision-makers require timely and accurate data to effectively formulate, implement and monitor policiesor investments. However, data are often unavailable or too outdated to be effectively used in

the policymakingprocess. This has increased the appeal of collecting data with computer assisted personal interview(CAPI) methods, which significantly improve the timeliness and quality of data. During July-August 2015, the Ministry of Livestock and Fisheries Development (MLFD), with the technicalassistance of the Global Strategy to Improve and Agricultural Statistics (GSARS), carried out a baseline surveyin Tanzania using the CAPI software called Survey Solutions. This was the first time that the MLFD usedCAPI and the endeavor was followed closely by other government institutions considering wider applicationof CAPI. This paper reports on the application of CAPI methods byMLFD, including training requirements,allocating skill sets across teams, optimizing logistics, monitoring field work, cost, challenges, and lessonslearned. The primary conclusion is that using Survey Solutions is an appropriate method for collecting data in developingcountries especially when rapid appraisals are needed. The initial cost of using Survey Solutions wasoffset by increased data quality, reduced time between data collection and analysis, and savings generatedfrom not having to print and transport questionaires. Furthermore, certain tools such as SIM cards frommore than one company, and additional training on tablet configuration for field supervisors can mitigateproblems like unreliable connectivity present in many developing countries. Finally, daily export, automatedtabulation, and reporting on paradata can improve the overall efficiency of the data collection process.

Using Mobile Devices for CAPI Data Quality Control in Developing Countries

Carol Place, *RTI International*
Kathy Mason, *RTI International*
Matthew (Clark) Letterman, *RTI International*

In recent years RTI has leveraged the increasing penetration of inexpensive and sophisticated mobile devices (smartphones and tablets) to improve quality control procedures for face-to-face surveys in developing countries. The proliferation of affordable high-quality hardware, along with increased access to internet connectivity has created the expectation among research funders that data collection will be carried out on mobile platforms. RTI conducts CAPI surveys using specifically tailored in-house data collection software that integrates three important aspects of quality control: 1) Collection of passive GPS data throughout the life-cycle of the interview 2) Recording of comprehensive paradata, such as time take to complete each question and interviewer movement within the instrument, to assess the quality of the interview administration 3) Computer assisted recordings of all or part of each interview (CARI) Our presentation will discuss RTI's practical experience using these features for quality control purposes in a diverse set of developing and middle income countries across multiple continents including Bangladesh, India, Nigeria, Uganda, Brazil, and Guatemala. In addition to discussing the technical aspects of the quality control mechanisms, this paper will also focus on issues such as integrating new technological approaches with local data collection partners who have varying levels of experience with CAPI surveys and who work in multiple languages. We will also discuss future uses of technology to increase the quality of CAPI survey data collected in developing countries.

Harmonizing CAPI Data Collection for a Youth Survey in Four African Countries

Sarah Hughes, *Mathematica Policy Research*
Matt Sloan, *Mathematica Policy Research*
Anca Dumitrescu, *Mathematica Policy Research*

Conducting surveys in resource-poor environments in sub-Saharan Africa (SSA) poses challenges for collecting high quality data. A rapid transition from paper and pencil personal interviews to computer-assisted personal interviews conducted on smartphones, tablets or laptops in many SSA countries holds

promise for improving the data quality and the efficiency with which the data are delivered to researchers and survey sponsors. When conducting multi-country studies, however, researchers must manage a variety of logistical challenges, including, but not limited to, institutional and cultural contexts that affect questionnaire design, differential training and experience of data collection teams, and limited opportunities for field observation within constrained budgets. In this paper, we describe an effort aimed at harmonizing our approach to CAPI data collection across four countries in sub-Saharan Africa. The multi-country longitudinal survey supports an impact evaluation of The MasterCard Foundation's large-scale secondary scholarship program in Ghana, Ethiopia, Rwanda and Uganda. We describe the steps taken in converting four separately-programmed questionnaires into a single multi-country instrument and the processes used for training field teams, implementing field operations and overseeing data collection. We will focus particularly on the collection and use of paradata to conduct day-to-day oversight of sample flow, field team actions and response trends from a distance.

Survey Planning and Logistics in Africa: Catching Up and Setting the Bar High in the 21st Century
Mari Harris, *Ipsos Public Affairs*

For many years, all surveys in Africa have been done via the rather long, cumbersome and laborious route of pen-and-paper data collection (PAPI), expensive travel to remote areas of the continent as well as the implementation of complicated selection procedures and physical data entry. The error rates were high, data cleaning was a special art and there was very little that could be done if - according to the selection procedure - the 'wrong' person was interviewed in a remote village 800 miles away. However, CAPI (Computer Assisted Personal Interviewing) and the exponential growth of mobile technology have changed all these things – for the better! For instance, the triple random selection procedure of choosing a stand, household on the stand and individual in the household that could have taken up to half an hour in the past is now done in a few minutes – this will be demonstrated in the presentation. All these changes were contributing to a total revolution and not just a small storm in a teacup – this paper will highlight the evolution of this process and will illustrate that, although survey planning and logistics in Africa still warrant detailed attention, the continent is truly at the forefront of benefiting from technological developments. Africa is indeed setting the bar of survey research high and it deserves this position: the continent is home to about a billion people, many countries in Africa record very high economic growth rates (especially when compared to the so-called "First World"), literacy rates are improving and more girls are attending school. We still have a long way to go, but it is important to take stock of how far we have come as we prepare for the next growth spurt.

Challenges for Random Probability Sampling in Africa
Virginia Nkwanzi, *Ipsos Public Affairs*

Research in African has evolved in the past few years evidenced by an increased appreciation for the science. Increasingly more organizations are using research findings to make decisions. Methodologies designed in developed economies do not always yield similar results in "non-developed economies", especially so in Africa. In Africa, one of the main challenges facing researchers has been in the area of sampling. This challenge differs substantially among countries and regions, affecting the application of sampling methodologies. This paper highlights challenges of sampling in Africa and possible, suggestions documented. Highlighting these issues forms a platform for experience sharing and it is hoped that this information will assist in research planning. These challenges arise at the planning and implementation phases of research studies. Census data, translations and effects of artificial and natural phenomena are some of the challenges affecting the planning process. Those occurring during implementation cover

women & culture, participation, close knit societies, political sensitivities and age variables. These will be discussed in detail in the paper.

Improving the Quality of Online Surveys

Repeat After Me: Effect Replication Across Online Nonprobability Sample Sources

Nicole R. Buttermore, *GfK Custom Research*
Frances M. Barlas, *GfK Custom Research*
Randall K. Thomas, *GfK Custom Research*

With the growing interest in online non-probability samples, one emerging concern in survey science is the ability to replicate effects, such as experimental results and stable brand ratings, across different sample sources. Foundations of Quality 2, a major study investigating survey quality conducted in 2013 by the Advertising Research Foundation, collected data from 17 different non-probability sample providers with each providing three separate samples of about 1,100 respondents. For this research, we examined four major content areas for replicability of findings: an experiment on response formats for past six month purchase, an experiment on new product evaluation using different response formats, evaluations of 27 different brands, and political spending priorities. To assess replicability, there were two major measures for each content area: bias (i.e., average absolute divergence from the overall mean across samples) and the relative ordering of results (i.e., high ratings for the same brand across samples compared to low ratings for another brand). We used demographically weighted data for comparisons to eliminate confounding factors due to profile differences across sample providers. For the response format and new product evaluation experiments, we found fairly high divergence across sample sources, but the relative rates of endorsement and new product preferences remained fairly constant. For the brand ratings, we found much greater absolute divergence among brands rated in the middle and less divergence at the endpoints, but the overall ordering remained fairly constant across samples. Finally, for political priorities, we found that for 15 of the 17 providers there was a consistent ordering, but two had divergent orderings of priorities. We discuss these results in terms of bias across non-probability samples and what we can generalize across samples.

Calculating Standard Errors for Nonprobability Samples when Matching to Probability Samples

Adam Lee, *ICF International*
Randy ZuWallack, *ICF International*

Integrating a non-probability panel survey with a probability based survey using statistical matching is a promising model for leveraging the statistical benefits of a rigorous probability sample with the cost and timeliness benefits of non-probability web. Statistical matching, or data fusion, is the process of combining data from two separate sources by linking members with similar characteristics. The traditional statistical matching challenge is to join two disparate sets of variables (Y, Z) in order to create a synthetic record jointly measuring Y and Z. In a nonprobability to probability matching application, the focus may be less on joint distributions and more on weighting. The probability sample, which

represents the population, provides the distribution to calibrate the nonprobability sample. Each person selected in the probability sample is assigned a statistical match from the nonprobability sample and inherits the nonprobability data from that match. The questions we explore in this presentation is—does this statistical matching approach preserve the sampling properties of the probability sampling, allowing the calculation of sampling errors and subsequently inferential statistics? Are there other forms of variability that must be accounted for in order to capture the total variance? Is the variability from the web panel a good estimate of the population variance? Our research uses data from the National Health Interview Survey, the National Alcohol Survey (dual-frame RDD) and a non-probability web survey.

Assessing the Accuracy of 51 Nonprobability Online Panels and River Samples: A Study of the Advertising Research Foundation 2013 Online Panel Comparison Experiment

Yongwei Yang, *Google*
Mario Callegaro, *Google*
Kathrine Chin, *Buros Center for Testing*
Ana Villar, *City University London*
Jon A. Krosnick, *Stanford University*

Survey research is increasingly conducted using online panels and river samples. With a large number of data suppliers available, data purchasers need to understand the accuracy of the data being provided and whether probability sampling continues to yield more accurate measurements of populations. This paper evaluates the accuracy of a probability sample and non-probability survey samples that were created using various different quota sampling strategies and sample sources (panel versus river samples) on the accuracy of estimates. Data collection was organized by the Advertising Research Foundation (ARF) in 2013. We compare estimates from 45 U.S. online panels of non-probability samples, 6 river samples, and one RDD telephone sample to high-quality benchmarks -- population estimates obtained from large-scale face-to-face surveys of probability samples with extremely high response rates (e.g., ACS, NHIS, and NHANES). The non-probability samples were supplied by 17 major U.S. providers. Online respondents were directed to a third party website where the same questionnaire was administered. The online samples were created using three quota methods: (A) age and gender within regions; (B) Method A plus race/ethnicity; and (C) Method B plus education. Mean questionnaire completion time was 26 minutes, and the average sample size was 1,118. Comparisons are made using unweighted and weighted data, with different weighting strategies of increasing complexity. Accuracy is evaluated using the absolute average error method, where the percentage of respondents who chose the modal category in the benchmark survey is compared to the corresponding percentage in each sample. The study illustrates the need for methodological rigor when evaluating the performance of survey samples.

Bias Reduction Through Rural Coverage for the AmeriSpeak Panel

Steven Pedlow, *NORC at the University of Chicago*
Jie Zhao, *NORC at the Unversity of Chicago*

The NORC AmeriSpeak Panel is a household, multi-client panel to support NORC's mission to deliver reliable data and rigorous analysis to guide critical programmatic, business, and policy decisions. AmeriSpeak uses the NORC 2010 National Frame to draw an address-based nationally representative sample of all Americans with full urban and rural coverage. The rural coverage provided by the NORC 2010 National Frame allows the NORC AmeriSpeak Panel to have the most coverage of the U.S.

population among multi-client panels in the US. While there is great difficulty in determining the coverage achieved by opt-in non-probability panels and by phone surveys, the US postal address database used for address-based sampling (ABS) provides city-style addresses for 92 percent of the U.S. population. This excludes post office boxes and rural route addresses that are mailable, but not reachable through in-person or telephone modes. In areas not well covered by the postal address database, the NORC 2010 National Frame enhances the ABS frame with in-person listing. We estimate that this increases the coverage by five percent. We also attempt to reduce bias by including these households in rural areas that would be missed by ABS studies. For the 2014 AmeriSpeak pilot, 17 out of 342 recruited households (5.0 percent) might have been missed by an ABS-only method. Even with this small sample, bias reduction was observed in education, single-person and two-person households, internet access, and marital status. Bias reduction was not observed in age and telephone preference. This paper will present whether these patterns are also present in the much larger main sample of approximately 7,500 households.

Grids and Online Surveys: Do More Complex Grids Induce Survey Satisficing? Evidence from the Gallup Panel
Mengyang Wang, *University of Nebraska–Lincoln*
Allan L. McCutcheon, *University of Nebraska–Lincoln*

High satisficing behaviors and item non-response can be greatly detrimental to survey quality; therefore online surveys must be designed with great care in order to minimize satisficing, increase responses and in turn to maximize quality. One derivative design is grid-style questions. However, this question type represents a double-edged sword to online survey market. On the one hand, the benefits that grid questions bring to the survey industry is to reduce time-consuming formats and shorten the length of questionnaire. On the other hand, there are several challenges that grid formats pose to measurement quality, including higher item non-response and dropout rates as well as an increase in measurement error. Operationally, grid questions can have very different levels of complexity which presents many challenges when using these types of questions in surveys. This research focuses on an experiment designed to explore data quality difference when responding to grid questions depending on: 1) the objective complexity; 2) the position of the grid; and 3) the interaction between grid's position and complexity. This study will allow us to evaluate the paradata, in combination with substantive data, from a fully randomized experiment conducted on the web component of the Gallup Panel. To increase the power of this analysis from previous studies, each respondent was assigned a questionnaire including a fully randomized ordering of three grid questions of varying complexity between the beginning, middle and end of the survey. Evidence from preliminary analyses indicates that grids with higher complexity are more likely to induce data quality problems, which demonstrated prominent measurement error among panel respondents. Moreover, as the survey progresses, the effects of grid complexity on data quality carried over to later questions, indicating respondents may be fatigued and thus become more likely to be satisficing on more complex grids at the end of the survey.

Issues in Media & Public Opinion

Is Social Media the New Forum for Discussions of Online Articles?: A Comparison of Reader Comments on Social Media and in Website-based Comment Sections

Ashley M. Schaad, *ICF International*

Online comments are increasingly being phased out by online news sites and magazines, citing factors such as the rise of social media, an increase in online trolls, and the poor quality of online comments (Gross, 2014). Many of these sites urge their readers to move their discussions to social media instead. However, there are several factors that make the two less than comparable, including the lack of a central location for the discussion and the more private nature of many social media sites. This paper presents a comparative analysis of reader comments posted in the online comments section of several news articles and reader comments to the same articles posted on the social media sites Twitter and Facebook. The articles used for this analysis come from The Washington Post, as well as the military publications The Army Times and The Marine Corps Times and focus of the contentious topic of the U.S. military's recent gender integration efforts.To mimic the quality of discussion available to members of the public wishing to discuss the articles online, only publicly-available comments were reviewed.

Does the Nightly News Still Matter?: Measuring the Effectiveness of Traditional Media to Mold the Opinion of Influential Americans

Orin T. Puniello, *Ketchum Global Research & Analytics*
Marni Zapakin, *Ketchum Global Research & Analytics*
Rebecca Strauss, *Ketchum Global Research & Analytics*
Patrick O'Donnell, *Ketchum Global Research & Analytics*

Communications strategies traditionally revolved around getting your message into the nightly network news and daily newspapers. However, the supremacy of traditional media has been challenged as an effective means of messaging with the rise of alternative platforms. Traditional media faces challenges from the internet, social media, and changing consumer patterns. The net effect of these challenges is that the media landscape has been decentralized, and now there is almost infinite ways for consumers to get information. This decentralization leads us to question traditional media as an effective tool for messaging? Put another way, given a world of finite resources, is a focus on traditional media still a smart communications strategy? We focus on the effectiveness of traditional media to shift opinions of "Influential Americans" which are defined as people that participate in three or more activities that are proven indicators of civic engagement. We analyze this influential group, which is approximately 10%-15% of the general public, because they been proven to lead and influence trends and policies. We examine this question by examining a 2015 survey of 1400 "influentials" for one of the largest healthcare institutions in the United States. The survey is part of an ongoing effort by this institution to optimize its communication strategy to increases its reputation, increase being seen as a leader in the healthcare field, and increase willingness to travel to the institution for serious health conditions.   Using a series of logit and OLS regressions models we isolate the effects of exposure to different types of traditional and non-traditional media channels has on influencing influential American's perceptions and opinions. We find that exposure to traditional media is still effective at influencing opinions, and often performs better than social media.

Values Priming and Press Performance: How Media Crises Activate Latent Attitudes and Shape News Evaluations

Erik P. Bucy, *Texas Tech University*
Paul D'Angelo, *The College of New Jersey*

Media researchers for decades have been interested in understanding and measuring the dimensions that individuals apply in evaluating the credibility of news. But aside from tracking the vicissitudes and credibility ratings of each new platform, the area has received little conceptual development. We address this deficit by theorizing news evaluation within a general framework of priming research, specifying a dynamic process of press-priming that plays out when a media scandal, press crisis, or other noteworthy event generates metacoverage that brings press values and media attitudes into play. Our model hinges on two important insights. First, news performance evaluations should be situationally defined in relation to the transgressive events and media missteps in which credibility emerges as a salient consideration. Second, the enduring values and attitudes that citizens develop towards media are key moderators of news assessments like credibility and trust. In this paper we identify accountability coverage—stories about the press that criticize news organizations and journalists for overstepping or violating professional standards and ethical norms—as the content mechanism that activates latent attitudes or normative expectations about press behavior. The model focuses on two core values: press freedom and media responsibility, which we operationalize as scale items into a typology of four different media outlooks. Press-priming effects are hypothesized to occur among those whose media attitudes are activated in the processing of paying close attention to scandal coverage (Bucy, D'Angelo, & Bauer, 2014). We thus sketch an integrative theory of news evaluation while explicating real-world relevance by using survey responses to recent media scandals in the US (Brian Williams' resume stretching; Rolling Stone's misreporting of an alleged rape at U. Virginia) and UK (the Jimmy Savile scandal at the BBC, phone hacking in the tabloid press) to demonstrate the utility of the model in predicting media perceptions.

Methodological Brief: Measurement Error & Data Quality

Satisficing vs. Optimizing in the Survey Response: An Evaluation of a Concept

Duane F. Alwin, *Pennsylvania State University*

The concept of "satisficing" vs. "optimizing" has gained considerable popularity among survey methodologists as a way of explaining differences in response styles among survey respondents. The idea is that when cognitive and motivational demands of choosing optimal responses to survey questions are high, respondents often respond with merely satisfactory answers, rather than to optimize their response choices. In this paper I discuss the concept of "satisficing" vs. "optimizing" as it is applied to the survey response. I begin with a discussion of the origins of the concept in social psychology in the work of Herbert Simon and the documentation of the history of the concept since its introduction to the survey research community. Research reveals that the concept has been quite popular since its introduction to the survey research community at the annual meetings of AAPOR in Hershey, Pennsylvania in 1987 [see Krosnick & Alwin, GSS Methodological Report No. 46, March 1987]. This paper draws attention to the popularity of the concept in survey methodology and by reviewing more than 400 papers that employ this concept in understanding the nature of responses to questions in surveys. A content analysis of these applications of the concept reveals several useful conclusions: (1) the use of the concept is typically used in an ad hoc fashion, employed primarily to account for patterns; (2) there are few efforts to measure the degree of satisficing directly; and (3) few experimental results

exist that would definitively suggest satisficing strategies. The paper concludes with an evaluation of the concept, suggesting that the concept of "satisficing" vs. "optimizing" rests on a crude psychological concept, and that more theoretically-grounded work is needed in order to understand the strategies respondents use to deal with cognitive and motivational burdens of survey questions.

Analyzing Check-All-That-Apply Responses From Large-scale Surveys
Ayrin C. Molefe, *American Institutes for Research*

Although researchers often encounter survey questions that allow respondents to choose more than one response option, the approaches available to analyze responses to such questions are not commonly known, easy to understand, or simple to implement. This current state of affairs is likely due to the fairly recent development of methods for analyzing check-all-that-apply (or multi-response) questions and their general absence in the syllabi for research methods or even advanced statistics courses. A typical research question of interest with this type of data is whether the response patterns are the same across two or more independent groups of respondents (e.g., males and females). Because a respondent can be classified into more than one response category, the usual statistical tests (e.g., Pearson's chi-square test of independence) are not applicable. When such data come from large-scale surveys, the analytic methods are even more challenging because one has to take into account the complex sampling design. The proposed paper will extend and illustrate the method of Katz & McSweeney (1984) to analyzing multiple response data from surveys with complex sampling designs. One such survey is the Education Longitudinal Study of 2002 (ELS: 2002) which asked, among other questions, high school sophomores in 2002 about their postsecondary educational aspirations. Respondents who indicated that they did not aspire to go beyond high school were asked to check all options that apply from eight possible reasons for not aspiring to go to college. Using these data, the proposed paper will demonstrate how to assess whether the reasons selected differed between rural and nonrural students. The paper will also describe a procedure for conducting post-hoc comparisons. Given the ubiquity of surveys with check-all-that-apply questions, this paper will be a useful resource for researchers who need to draw sound conclusions based on multi-response data.

Data Quality and Presentation Choice in Online Surveys: Evaluating the Potential of MaxDiff and Fractional Factorial Design
McKenzie Young, *Global Strategy Group*
Andrew Baumann, *Global Strategy Group*
Lyndsey Collins, *Global Strategy Group*
Damla Ergun, *Global Strategy Group*
Michael Smith, *Global Strategy Group*

The effect of question format on data quality is a hotly debated area of survey research. Past studies find that a range of factors including the type, labeling, and number of response options significantly affect how respondents interpret and answer questions. The differences between standard attitudinal question types like open- versus closed-ended and rating scales versus ranking questions in terms of data quality are well-established, yet less is known about how newer approaches like MaxDiff and fractional factorial designs compare to these methods on data quality. For example, rating scales are considered superior to ranking questions when respondents are asked to evaluate multiple items because ranking questions require significantly more effort from respondents (Alwin & Krosnick, 1985). While ratings are cognitively easier for respondents, they often suffer from non-differentiation, the process by which respondents give identical answers to multiple items. Approaches like MaxDiff aim to

overcome these shortcomings by presenting respondents with a subset of the full set of options through forced choice questions which eases the cognitive burden and allows researchers to take advantage of the superior reliability and validity associated with rankings. A MaxDiff approach provides a unique rank for each item based on respondents' answers to every combination of pairs of messages, and the Fractional Factorial design presents each item to exactly half of the sample, simulating a random independent sample for each message. This national online study, conducted specifically to examine differences in data quality between several attitudinal measures including rating scales, a MaxDiff design, and a fractional factorial design, provides much needed input into current debates in public opinion research on item and format choice and their effect on response quality.

Using a Calendar vs. Frequency Grid in a 2014 National Mail Survey

Danielle Battle, *American Institutes for Research*
Rebecca Medway, *American Institutes for Research*
Meghan McQuiggan, *American Institutes for Research*

Reducing the burden placed on respondents is necessary when asking them to recall events. Item design, especially in self-administered questionnaires, can reduce burden by providing cues that aid in recall. There are two different approaches that have been used to collect this type of data. One is a calendar format, in which respondents are asked to mark the days of the week that an event occurs. One challenge of this method is possibly higher levels of straight-lining due to the complexity of the response process. The second approach is a traditional frequency grid, in which the response options are number of days per week (e.g. once a week, 2-3 days per week, etc.). However, this method might lead to overreporting due to respondents increasing their reported frequency to allow for forgotten events. Overall, some research has suggested that calendar formats may enhance recall and improve data quality, but methodological research comparing the two approaches has been minimal. As part of the 2014 National Household Education Surveys Feasibility Study, a self-administered mail-based survey, a split panel experiment was conducted to determine whether a weekly calendar format or an event frequency format would be preferable for future administrations. Half of the sample received a form with the items in a calendar format, and the other half received a form with the same questions in a frequency format. The proposed analysis of the calendar format (n=2,850) versus the frequency format (n=2,870) will examine potential recall issues such as over- or underreporting by comparing response distributions. Initial findings indicate there is a statistically significant association between the item format and the distribution of responses. The analysis will also look at data quality by comparing unit response rates, item non-response rates, and straight-lining patterns to determine which format is best suited for a self-administered questionnaire.

Are Reference Periods Referenced?

Jennifer Edgar, *U.S. Bureau of Labor Statistics*
Gina Shkodriani, *Westat*

For many survey concepts the reference period is a critical part of a question, directing the respondent what point in time to consider when arriving at their answer. For household respondents, this is most often a mental exercise, as recall information from the specified period. For establishment surveys however, respondents often have to consult records to arrive at an answer. The extent that the reference period in the question matches the available data (e.g. monthly payroll reports) likely determines how a respondent uses the reference period in their repose process, if at all. BLS contracted with Westat to conduct 100 exploratory interviews exploring the feasibility of using an alternative reference period for the Job Openings and Labor Turnover Survey (JOLTS). This presentation addresses

two main questions: Does the reference period align with establishments' records or data systems? Do establishments actually report data for the specified period, not just the data that is easiest to access? Findings highlighted the importance of establishments being able to align the data request with their records. Some respondents could easily adjust their data systems while others could not. Additionally, it was found that most establishments reported data for the specified time period, if instructions were clearly communicated. However, some participants, even in this research project where they had volunteered to participate, reported the most accessible data, regardless of reference period. This presentation will share the results from this study, highlighting the implications for survey designers working on both household and establishment surveys. We provide specific suggestions on how to tailor instructions to increase the likelihood that the reference period is understood, a prerequisite for its use.

The Value of Education: Alternatives Using Educational Attainment for Weighting
Kristie Healey, *ICF International*
Randy ZuWallack, *ICF International*
Brian Orleans, *ICF International*

We regularly calibrate our sample surveys to match the population based on a number of demographic variables--age, gender, race/ethnicity, marital status, and educational attainment. As survey nonresponse increases, particularly differential nonresponse between population subgroups, these weighting adjustments have become particularly important to mitigate the risk of bias in the survey estimates. Throughout the early 2000s, increasing differential nonresponse by age group resulted in samples that were heavily skewed toward older adults and underrepresented adults under the age of 35. However, as we shifted toward dual-frame landline and cell phone samples, the age distribution gaps were reduced. Similarly our race/ethnicity distributions benefited from the addition of cell phone sample. However, adding cell phones did nothing to curb the growing differential in educational attainment in telephone samples. Examining the Behavioral Risk Factor Surveillance Study (BRFSS) from 2001-2014, the percentage of adults with a college education before demographic weighting has increased from 28.7% to 37.6%, over-representing the 27.5% of adults with a college education (2014 ACS). In 2014, the education weighting variable ranked first in terms of the highest between class variability in the weights among all demographic weighting variables for 32 states and second for 19 states. In this presentation, we explore the use of education as a weighting variable using data from the BRFSS and the National Adult Tobacco Survey. The issues we study include which education levels should be grouped together (e.g. some college and associates degrees) as well as whether there are other weighting variables that interact with education, including age and gender.

Methodological Challenges and Opportunities in Web Survey Usability Evaluation

Panel organizer: Lin Wang, *U.S. Census Bureau*

At the U.S. Census Bureau, Web surveys have become a major data collection mode. In order to minimize the potential measurement and non-response errors introduced during the self-administration of Web survey, usability evaluation is crucial. Web survey usability evaluation, as a discipline of scientific investigation, requires sound methodology. Yet, Web survey usability evaluation is a relatively young field that is filled with challenges and opportunities. In this panel, we will introduce the general approach, practiced at the U.S. Census Bureau, to Web survey usability evaluation, followed by discussing five major challenges that we had experienced in our practice. The five challenges are (1)

conducting cognitive probing techniques, (2) utilizing and interpreting eye tracking data, (3) evaluating accessibility on mobile device, (4) incorporating usability evaluation in the agile software development process, and (5) comparatively assessing usability of desk-top-based and mobile-device-based Web surveys.

Cognitive Probing Methods in Usability Testing – Pros and Cons
Elizabeth Nichols, *U.S. Census Bureau*

Think-aloud is one cognitive probing technique often used in usability testing. When think-aloud is applied during the usability testing, the participant is instructed to give a running commentary while performing a usability task. The participant's verbal commentary can include such things as what he or she is looking at, reading, thinking, feeling, and doing to answer the questions. Collecting this type of qualitative data is useful because it helps find the root causes of usability problems. This technique is modeled after Ericsson and Simon's protocol analysis. The technique used in usability testing is very similar to the probing techniques used in cognitive testing which include concurrent and retrospective probing. Cognitive testing is conducted to pinpoint problems in question wording and response options. These two types of probing often overlap in online surveys. Depending upon the purpose of the usability evaluation, we have experimented with a variety of probing methods in usability testing, including retrospective probing, think-aloud, and think-aloud with concurrent probing. This talk will discuss these techniques and their trade-offs in light of the usability evaluation purpose.

Association of Eye Tracking with Other Usability Metrics
Erica Olmsted-Hawala, *U.S. Census Bureau*

Eye tracking data are routinely collected during Web survey usability testing at U.S. Census Bureau. While oculomotor activities are informative for inferring respondent performance of completing Web survey, challenges remain to establish performance indicators using eye tracking data. This talk will highlight some of the challenges faced when incorporating eye tracking into usability studies of electronic surveys; and will explore interpreting results with respect to the other common usability metrics of effectiveness, efficiency, and satisfaction. Examples of how results can support findings will be presented. In addition, the talk will touch on both the technical and the methodological challenges that may be encountered when using eye-tracking technology on mobile devices.

Making Mobile Web Surveys Accessible
Lawrence Malakhoff, *U.S. Census Bureau*

The proportion of Web access is increasing for mobile devices but declining for the desktop computer. Accordingly, designers of survey instruments should consider making their Web surveys conform to Section 508 guidelines not only for the desktop, but also for mobile devices. Persons with low or no vision navigate on a desktop computer by tab, enter and arrow keys with a screen-reader speaking labels, instructions, and response options. Mobile device keyboards cannot be used to navigate this way through any Web survey instrument. Instead, persons using mobile devices use gestures to navigate. When a screen-reader user moves their finger on the screen, a correctly programmed interface will speak the instruction or question text, button labels, and status of a radio button or checkbox response option when touched. The user will then hear this feedback and can provide a response. This presentation will discuss guidelines for creating an accessible mobile Web survey using Apple Siri and Google Talkback.

Usability Testing within Agile Process
Temika Holland, *U.S. Census Bureau*

Incorporating usability testing into the development cycle of mobile survey applications is a very important approach to human-centered design. The U.S Census Bureau is implementing the agile software development methods for its Web survey instruments and tools. The agile practice provides us with both challenges and opportunities for usability testing. On one hand, it allows for usability feedback and recommendations to be considered in the early iterative phase of development rather than at the end where changes may be difficult to implement. On the other hand, the amount of time that can be afford for conducting usability testing is getting even less. In this talk, we will use the project, the Census Operations Mobile Platform for Adaptive Services and Solutions (COMPASS), as an example to illustrate the role of usability testing during the agile development cycle, obstacles or challenges, and impact on the iterative design of Web-survey applications.

Exploring Why Web Surveys Take Longer to Complete on Smartphones than PCs: Findings from a Within-subjects Experiment
Christopher Antoun, *U.S. Census Bureau*
Alexandru Cernat, *University of Essex*

As respondents increasingly respond to Web surveys on their smartphones instead of their personal computers (PCs), one finds that respondents take longer to complete Web surveys when using phones than PCs, and longer completion times may increase the risk of breakoff, but little is known about what drives these differences. In the current study, we compare response times for 895 respondents who completed the same Web survey once using a phone and once using PC as part of a methodological experiment in the LISS panel (Longitudinal Internet Studies for the Social Sciences), a Dutch probability-based Web panel. We analyzed survey-level times, item-level times, self-reported information about where respondents completed the survey and whether they were multitasking or distracted at the time. Linear mixed models are used to account for paired observations and questions nested within respondents. We focus on three potential factors: item characteristics, respondent characteristics, and respondent context. We expect that longer and more complex items will have a larger effect on mobile Web completion times. We also expect that respondents with less familiarly or comfort with smartphones, and those who are older, will be most responsible for longer response times in mobile Web because of their potentially decreased visual-motor coordination and lower levels of familiarity with mobile browsers. Finally, we expect that distracting contexts should inflate response times, and responding in these contexts is more likely when using smartphones. The results will help shed light on the device differences in completion times and inform strategies to reduce such differences.

Question Order, Question Context and Questionnaire Length

Satisfied or Dissatisfied? Does Order Matter?
Jolene D. Smyth, *University of Nebraska-Lincoln*
Richard Hull, *University of Nebraska-Lincoln*

Many surveys use satisfaction/dissatisfaction questions with Likert scales. Best-practices design guidelines often suggest balancing the question stem by mentioning both satisfaction and dissatisfaction and keeping the response options in order. However, little research has examined the impact of the order of the constructs of "satisfied" and "dissatisfied" in the question stem and response options. That

is, should "satisfied" or "dissatisfied" be mentioned first in the question stem? Which should be mentioned first in the response options? Does it matter if the order in the stem does not match the order in the response options? This paper reports the results of an experiment examining these issues. The experiment was embedded in the 2015 Nebraska Annual Social Indicators Survey (NASIS). 3,500 addresses sampled from the DSF were randomly assigned to one of four versions of the questionnaire (i.e., a 2x2 design) in which the order of the terms "satisfied" and "dissatisfied" were switched in the question stem and response options (with the ordinality of the response options always maintained) for three questions. Data were collected between August and October 2015 and responses were received from 1,143 adults selected from households using the next birthday method (AAPOR RR1 32.7%). Preliminary analyses show that the response distributions were significantly affected by the experimental treatments. Further analyses will examine what experimental factors are leading to these differences as well as whether the experimental treatments impacted item nonresponse rates for these items. The paper will conclude with discussion of implications for the design of satisfied/dissatisfied questions in mail surveys.

Instructions in Self-administered Survey Questions: Do They Improve Data Quality or Just Make the Questionnaire Longer?

Cleo Redline, *National Center for Education Statistics*
Andrew Zukerberg, *National Center for Education Statistics*
Chelsea Owens, *National Center for Education Statistics*
Amy Ho, *National Center for Education Statistics*

Pre-testing techniques utilized in the development of production self-administered questionnaires, such as cognitive interviewing, often identify items where respondents misinterpret or are unclear about the meaning of terms in a question. Typically, this finding results in a recommendation to add instructions to an item, which has the detrimental effect of lengthening the questionnaire. Previous experimental research has shown that instructions have an effect on the estimates when the instructions counter the way many people naturally tend to think about a concept. For example, an instruction to exclude sneakers from a count of shoes will reduce the estimate of shoes because many respondents tend to think of sneakers as shoes. In addition, previous research has shown that instructions placed before questions are more effective than those placed after. However, few studies have looked empirically at whether or not instructions that are the product of actual production pre-testing techniques are similarly effective or useful, and worth the extra length they create. Nor have many other factors been examined that might influence the effectiveness of instructions. To examine these issues further, we report on an experiment that was administered to a nationally representative sample by web. Production questions and instructions were selected from a national teacher survey. In addition, questions and instructions were intentionally created to counter teachers' natural conceptions of terms. These items were compared to a control group with no instructions. Utilizing a factorial experimental design, we also varied three factors that were predicted to alter the effectiveness of instructions: their location, format, and wording. Although the findings of this experiment are clearly generalizable to the web, arguably, these findings extend to mail surveys too.

Investigating Measurement Error through Survey Question Placement
Ashley R. Wilson, *RTI International*
Jennifer Wine, *RTI International*
Natasha Janson, *RTI International*
John Conzelmann, *RTI International*
Emilia Peytcheva, *RTI International*

Previous designs of a national, cross-sectional survey on how students and their families pay for postsecondary education have focused on measuring and reducing nonresponse bias. However, the potential for measurement error in the survey estimates remain unstudied. This paper presents the results of several experiments aimed at investigating measurement error in this comprehensive student survey. We look at two indicators of measurement error – differences in responses to questions as a function of their placement in the survey (early vs. late) and response accuracy as a function of question placement in the survey. For differences in responses as a function of question placement, we examine responses to fictitious student loan questions and reverse coded questions placed in different parts of the survey. In order to examine response accuracy, we compare responses to federal student loan questions placed early vs. late in the survey, with actual loan amounts available through an administrative data matching source. Overall, about 10% of respondents provided substantive responses to fictitious questions. Contrary to expectations, the percentage was significantly higher when the fictitious item was placed in the first half of the survey. In addition, a significantly higher percent of substantive responses was obtained when the interview was conducted by phone rather than self-administered (web or mobile web). Similarly, about 9% of respondents provided opposing opinions when answering a reverse-worded item later in the survey. Finally, respondents tended to overestimate their loan amounts regardless of where in the survey the item was placed – we found more reporting error to questions asked earlier rather than later in the survey, but the difference did not reach statistical significance. We explore the potential implications of these results for current and future survey designs.

Global Warming: The National and Global Public Opinion Challenges

Panel organizer: Mark A. Schulman, *Abt SRBI*

As climate experts warn about the dangers of global warming, elected officials around the world continue to grapple with the politics and economics of the issue. Clearly public opinion plays a key role in shaping the politics and national/international discourse on climate change. We have assembled some of the leading opinion researchers on climate change to address in-depth key substantive and methodological issues in the U.S. and globally. These issues include methodological/measurement issues, the dynamics of climate change opinion formation, the global landscape, where and how much are people concerned about climate change, what are the causes, the impact of the Pope's encyclical, and what effects of climate change trouble people most. Several of the presentations feature primary data.

Concern and Action: Public Opinion about Global Climate Change around the World

Katie Simmons, *Pew Research Center*
Jill Carle, *Pew Research Center*
Richard Wilke, *Pew Research Center*

As world leaders continue a decades-long struggle to address the threat of a warming planet, global climate change has solidified its place as a key international issue. While public opinion plays a key role in shaping national and international discourse on climate change, few cross-national studies explore public opinion on this issue in depth. Where and how much are people concerned about climate change? What effects of climate change trouble people most? And what should be done to address the issue? We will explore these questions using data from a spring 2015 Pew Research Center survey conducted in 40 nations from across the globe. The survey was conducted from March 25 to May 27, 2015 among 45,435 respondents. In 31 countries face-to-face interviews were conducted, and in 9 countries interviews were conducted via landline and cell phones. All samples were nationally representative. In this paper, we explore attitudes on a variety of issues related to global climate change, including concern about the issue, its immediacy and personal impact, the consequences of climate change, and potential solutions. We investigate differences between regions and countries, as well as within regions and countries. Additionally, we use multivariate analysis to evaluate the influence of demographic and attitudinal variables on concern about climate change and its effects.

Different Survey Organizations, Different Results: Do We Know What Public Opinion is on Global Warming?

Adina Abeles, *Stanford University*
Jon Krosnick, *Stanford University*
Bo MacInnis, *Stanford University*

During the last 25 years, three organizations have conducted many surveys measuring Americans' opinions about global warming via random digit dial telephone interviews (Gallup, the Pew Research Center, and Stanford University's Political Psychology Research Group (PPRG)) and have yielded many apparently contradictory findings about distributions of opinions cross-sectionally and about trends in opinions over time. This is disconcerting for numerous reasons, including its potential to reduce public confidence in polls and the impact of survey data on policy-making. This presentation will show that the reported differences do not reflect an inability for survey organizations to produce similar results when measuring the same construct. Instead, the differences appear to be the result of different question wordings. The talk will present graphs comparing the cross-sectional results and over-time trends produced by each question wording, highlighting striking differences between organizations. Then we will report results from experiments that randomly varied question wording while holding everything else about the data collection constant. The experiments asked wordings used by the three organizations and other wordings that 'walk' from one version of a question to another by gradually changing one word or phrase at a time. The results illustrate how survey findings are importantly contingent on even seemingly trivial aspects of question wording and suggest that the surveys done by Gallup, Pew, and PPRG are all valid but measure different opinions.

Can a Progressive Pope Sway Skeptics? Tracking Public Opinion on Pope Francis and Climate Change

R.B. Lull, *University of Pennsylvania*
H. Akin, *University of Pennsylvania*
N. Li, *University of Pennsylvania*
J. Hilgard, *University of Pennsylvania*
A.R. Landrum, *University of Pennsylvania*
K.R. Barnhart, *University of Pennsylvania*
K. Winneg, *University of Pennsylvania*
D.A. Scheufele, *University of Pennsylvania*
K.H. Jamieson, *University of Pennsylvania*

In the hyper-partisan U.S. political environment, where public attitudes toward climate change are sharply divided along party lines, commentators and pundits speculated about how public opinion would be affected by Pope Francis's visit to the U.S. Speculation was fueled by the political rhetoric that followed his encyclical Laudato si', an official Catholic document stating climate change is a moral issue demanding immediate global action. Previous research suggests two theoretical scenarios for how public opinion dynamics could play out in the aftermath of the encyclical and the Pope's visit: Pope Francis could leverage the Catholic Church's credibility on traditionally conservative issues, such as abortion, and compel skeptics to change their climate change attitudes; or his largely conservative constituency could maintain their climate change opinions and instead respond with negative perceptions of the Pope's credibility on this issue. In order to systematically track the effect of Pope Francis's climate change campaign on the opinions of U.S. Catholics and the broader population, we conducted large-scale phone surveys at four time points, ranging from before the encyclical release to after the Pope's visit. Our data include two nationally representative panel surveys with two waves each, complemented by two representative cross-sectional surveys. This allows us to capture opinion changes regarding climate change, other issues, and Pope Francis. We also asked respondents about religious and political beliefs and media use to track effects of media coverage of the encyclical and the U.S. visit. Analyses show that conservatives perceived the Pope to be less credible on climate change after the encyclical was released. Otherwise, the Pope's message resonated broadly and bolstered his favorability, even across religious denominations and among the non-religious. Analyses presented at the conference will report on the early rounds of data collection, as well as the post-visit panel waves currently in the field.

Finding Common Ground on Global Warming: Results from the October 2015 Yale-George Mason Survey

Geoff Feinberg, *Yale University*
Seth Rosenthall, *Yale University*
Anthony Leiserowitz, *Yale University*
Jennifer Benz, *NORC at the University of Chicago*
Trevor Tompson, *NORC at the University of Chicago*

People in the climate change community understand that talking with family and friends about global warming is a critical part of raising awareness and mobilizing action for effective climate policy. But discussing the subject with those who think global warming is not happening often lead to arguments over this highly politicized and divisive issue. There are, however, many points of agreement on climate-friendly policies, and productively discussing such issues can help move the discourse on climate

forward. For example, like other Americans, majorities of Americans who think global warming is not happening say they are extremely, very, or moderately concerned about air pollution (74%), loss of tropical rain forests (73%), loss of open spaces and undeveloped land (61%), extinction of plant and animal species (57%), the condition of the world's oceans (54%), and the fact that clean energy is not widely available (53%). Moreover, they either support or do not oppose the following: requiring drilling companies to publicly disclose all the chemicals they use when "fracking" (88%), funding research into renewable energy sources such as solar and wind power (70%), and regulating carbon dioxide emissions (62%). In our presentation, we will discuss these and other findings from the Yale AP-NORC Environment Poll, a survey conducted among the American public by the Yale School of Forestry and Environmental Studies and the Associated Press-NORC Center for Public Affairs Research late last year.

Climate Change: Opinion Formation and the 2016 Election
Gary Langer, *Langer Associates*

This paper will review current attitudes on climate change, the dynamics of opinion formation, and the likely impact of the issue on the 2016 presidential election. It will discuss related strands of research explored in ABC News/Washington Post polls and other surveys. These include the role of partisan predispositions in attitude formation, particularly in the presence of equivocal or contested facts; the phenomenon of expressed belief, meaning the propensity of respondents to telegraph their political or ideological preferences in answering policy questions; and skepticism about the effectiveness of government-led attempts to address difficult problems. The paper will suggest future research directions as well as review past results.

Advances in Adaptive Survey Designs: Developing a Bayesian Perspective

Panel organizer: Natalie Shlomo, *University of Manchester*

Adaptive survey designs employ different strategies or design features to different population strata. The strata are identified by auxiliary data from administrative data and/or from paradata. The employment of the strategies may take place in the form of interventions during data collection and/or through the optimization of design in between waves of the survey. The interventions and optimization are based on estimated design parameters such as stratum contact propensities, stratum participation propensities, stratum mode coverage propensities and stratum costs parameters. These parameters are estimated using a combination of expert knowledge, prior data and current data. As a consequence, the estimated parameters are subject to inaccuracy. Furthermore, the parameters will, generally, change gradually in time. A Bayesian approach towards interventions and optimization is natural as it is a tractable and convenient way to mix and weight expert knowledge, prior data and current data, to account for the resulting uncertainty in the design parameters, and to allow for gradual change. The Bayesian component to adaptive survey design consists of 1) prior distributions to parameters in models for nonresponse and measurement and 2) decision rules in interventions and optimization of design. The Bayesian Adaptive Survey Design Network (BADEN) funded by the Leverhulme trust in the United Kingdom consists of the universities of Manchester (network coordinator), Michigan and Southampton, the national statistical institutes of the Netherlands and Sweden, RTI international, and the US Census Bureau. For this panel proposal, BADEN aims to sponsor a session focusing on case studies with an emphasis on design features and monitoring to inform adaptive survey designs in a Bayesian framework.

A Bayesian Analysis of Mixed-mode Data Collection: Results from Four Case Studies

Lisette Bruin, *Statistics Netherlands*
Nino Mushkudiani, *Statistics Netherlands*
Barry Schouten, *Statistics Netherlands*

The most influential feature in survey design is the survey mode. Survey modes have a strong quality-cost differential and adaptation to population subgroups may, therefore, be very effective in terms of quality or costs. However, the survey mode impacts both nonresponse and measurement error, so that a combined view on both errors is needed. To date, all surveys at Statistics Netherlands are multi-mode. Designs are mostly sequential, starting with cheaper modes and allocating nonrespondents to more expensive interviewer-assisted modes. In an adaptive design, some population subgroups may receive a follow-up and other subgroups may skip the cheaper modes. We have monitored four general purpose multi-mode surveys in a Bayesian framework. Regression parameters in models for response, cost and mode effects are assigned prior distributions based on expert knowledge and historic survey data. The prior distributions have been updated on a frequent basis using observed survey data. Models contain the most relevant register variables linked to at the time of sampling as well as a few paradata variables linked during data collection. We highlight the results from the Bayesian analysis on the four surveys. We give specific attention to the population subgroups that are candidates for adaptation.

Modelling Length and Final Response Outcome of Call Sequences in the Swedish Labour Force Survey

Anton Johansson, *Statistics Sweden*
Peter Lundquist, *Statistics Sweden*
Sara Westling, *Statistics Sweden*
Gabrielle Durrant, *University of Southampton*

Nonresponse in sample surveys is an ever increasing problem with serious negative consequences to the quality of survey data. At the same time survey agencies face increased survey costs. Developing improved and more efficient data collection strategies is hence a pressing need for survey producers. This paper aims to reduce unproductive calls in the Swedish Labour Force Survey, in particular those calls that require many call attempts and interviewer working time but still end up in a nonresponse outcome. We use recent research (Durrant et al. 2015) to model response propensities and to identify good predictors for long unproductive call sequences. In particular, the aim is to identify such long and unsuccessful call outcomes early in the data collection process. Further strategies are also investigated to reduce the time spent on productive call sequences with the aim of reducing costs and increasing efficiency. Call records from the Swedish Labour Force Survey are investigated. To build informative nonresponse models the study uses rich data from both the Swedish register and paradata. The evaluation of models compares model predictions to the observed outcomes using already collected data. The research should be seen as a necessary step in building an environment for responsive and adaptive designs. The findings are expected to help survey agencies to reduce costs during data collection and to improve efficiency. [Durrant, G., Maslovskaya, O. and Smith, P. (2015) Modelling final outcome and length of call sequence to improve efficiency in interviewer call scheduling, Journal of Survey Statistics and Methodology (forthcoming)]

Interventions During Data Collection to Increase Response and Sample Representativeness: a Field Test Experiment and Responsive Design Simulation

Dan Pratt, *RTI International*
Jeffrey Rosen, *RTI International*
David Wilson, *RTI International*
Melissa Cominole, *RTI International*
Elizabeth Copello, *RTI International*
Andrey Peytchev, *University of Michigan*

Responsive design methods may be employed at multiple points during data collection to identify nonresponding sample members for targeted interventions to encourage participation with goals of increasing representativeness of the responding sample or reducing potential nonresponse bias in population estimates. Contributing factors to the success of such approaches include the determination of specific interventions which may encourage response and identification of particular nonresponding cases for targeted interventions which may improve sample representativeness if they participate. This presentation reports results from a field test of a national study conducted for the U.S. Department of Education's National Center for Education Statistics (the High School Longitudinal Study of 2009 [HSLS:09] Second Follow-up). A prior HSLS:09 data collection demonstrated the effectiveness of a responsive design model used during data collection to identify sample members who were underrepresented with regard to key survey variables. Likewise, interventions used among cases targeted based on model results also seemed effective in encouraging cooperation among targeted cases, although interventions were not evaluated experimentally. To assess the effectiveness of four specific interventions, a randomized design was implemented for the HSLS:09 second follow-up field test to allow for experimental evaluation. Due to the small field test sample size and because some interventions were under consideration for all main study cases (not solely cases targeted by responsive design methods), interventions were implemented based on random assignment to treatment conditions rather than a predetermined threshold from the responsive design model. We also analyzed field test results – both in terms of case selection effectiveness and intervention effectiveness – by simulating what may have happened if we targeted specific cases based on a responsive design model, rather than random assignment to experiment conditions.

Using Bayesian Methods to Estimate Response Propensity Models During Data Collection

James Wagner, *University of Michigan*

Response propensity models have been used to create inputs to adaptive survey designs. These inputs may be needed during data collection as triggers for design decisions. However, Wagner and Hubbard (2014) showed that estimates of response propensity models can be biased when fit on a daily basis during data collection using the incoming data. One solution to this problem might be to fix the estimated coefficients from these models using estimates from previous data collections. There are several disadvantages to such an approach. First, a suitably matched data collection must exist. A Bayesian approach allows for the incorporation of expert judgment into the model. Methods for eliciting priors from experts have been developed in the clinical trials setting. Further, using data from a previous data collection is not sensitive to any changes that may occur over time. Bayesian logistic regression models, with informative priors, may address these weaknesses. This paper presents results from an effort to develop appropriate priors for models fit on a daily basis during data collection.

Improving Response Propensity Estimation for Adaptive Design Interventions at the U.S. Census Bureau

Stephanie Coffey, *U.S. Census Bureau*
Benjamin M. Reist, *U.S. Census Bureau*

Interest in responsive and adaptive survey designs has increased in recent years as a framework for tailoring contact strategies to individual sample units. The specific tailoring strategy applied to units is selected in pursuit of a variety of data collection goals, including: improving representativeness within a fixed cost, maintaining representativeness while decreasing costs, or improving timeliness of progress. Whether tailoring occurs between or within a survey round, decisions are made based on existing information, which could include frame and sample data, paradata, and even response data, in the form of response rates, point estimates, or variances over time. Propensity scores are being used for adaptive design purposes at the Census Bureau. Balancing propensity models are used in the generation of R-indicators (Schouten, et al. 2009) to evaluate representativeness. Separately, response propensity scores are used to prioritize fieldwork. Both of these metrics, while aiming for different goals, mean that estimation of propensity scores during data collection is of high importance. Unfortunately, early in data collection, using only frame information and limited paradata collected on current cases, there is empirical evidence that response propensity scores can be biased and volatile. Poor estimates of propensity scores could lead to inappropriately selecting cases for adaptive design interventions. Rather than using frame information known for the cases in sample, and the limited paradata available, the Census Bureau is exploring using a Bayesian framework to estimate response propensities. Leveraging significant historical information within a given demographic survey as well as potentially across demographic surveys could improve estimates of propensity scores, allowing us to make better decisions during data collection. This presentation will cover situations where poor propensity model estimation could lead us to make sub-optimal data collection interventions, how we think a Bayesian framework will improve propensity score estimates, and issues we have encountered with estimation.

Measuring & Understanding LGBT & Gender Issues

Examining Public Support for Transgender Rights and Supportive Policies

Ashley Kirzinger, *The Henry J. Kaiser Family Foundation*
Jason Pierceson, *University of Illinois Springfield*

Recently, a great deal of attention has been paid to lesbian and gay rights, particularly the issue of marriage equality. While the transgender rights movement is often conflated with the lesbian and gay rights movement, transgender individuals face unique challenges that require a different policy response in many circumstances. Unfortunately, not much public opinion data exists gauging levels of support for transgender-supportive policies. In addition, very little is known on how much individuals know about issues surrounding transgender persons including awareness of specific terms and policies. In order to fill in this knowledge gap, we conducted an online survey using Qualtrics market research panels in order to establishing baseline measures of questions dealing with knowledge, awareness, and support for specific policies. This survey is important for two reasons. First, it is one of the first surveys asking a national audience about transgender issues. Second, it relies on a growing field of online survey methodologies in order to reach a national sample. The survey was conducted in July 2015 with 774 respondents including quotas for both gender and age. The sample also includes an oversample of individuals who identify as either lesbian, gay, bisexual, or transgender (75 total respondents). Using the weighted results, we find that personally knowing a person who is transgender, gender non-conforming,

or transitioning is a positive predictor of having a positive overall attitude towards transgender persons as well as being supportive of transgender inclusive policies in healthcare, K-12 education, prisoner's rights, and the U.S. military. (similar to 2009 Gallup findings and 2013 Pew findings regarding attitudes towards gay and lesbian individuals and policies). The results of this survey add to the growing discussion of national attitudes towards transgender supportive policies and whether methodologists can rely on a recruited online sample to survey a national audience.

Asking About Gender Identity in Surveys
Frances M. Barlas, *GfK Custom Research*
Nicole R. Buttermore, *GfK Custom Research*
Mansour Fahimi, *GfK Custom Research*
Randall K. Thomas, *GfK Custom Research*
Maya Grosul, *Gfk Custom Research*

Gender identity is an important indicator in research investigations across an array of substantive topics, such as health and employment, yet few national surveys include questions on gender identity. There is a growing body of research testing various methods of asking gender identity for inclusion in surveys. KnowledgePanel, GfK's online probability-based panel, is one of the first national surveys that asked questions about gender identity as part of its profile surveys since 2010. Initially, we asked respondents a yes/no question about whether or not they identified as transgender. Results showed a likely overestimate of the transgender population due to comprehension problems, particularly among Spanish survey takers. We subsequently conducted a series of surveys to test alternative ways of asking about gender identity, most recently testing a two-step method for assessing gender identity proposed by the Center of Excellence for Transgender Health at University of California – San Francisco in which respondents are asked their sex assigned at birth and their current gender identity. A sample of 2,874 KnowledgePanel members completed a survey with random rotation of the order in which these two questions were asked, as well as randomly varying response categories for the question on current gender identity. Respondents were also asked a series of follow-up questions to gauge their comprehension and comfort with the gender identity questions, as well as the extent to which they felt the survey questions allowed them to accurately tell us their gender identity. Results indicate that the new two-step gender identity questions produced an estimate of the transgender population that is more in line with best national estimates. Respondents also found the new questions easy to understand and answer.

Changing Public Opinion Towards LGBT Rights Between 1992 and 2012: A Comparison of Adults in Military and Civilian Sectors
Jacob P. Absalon, *University of Nebraska-Lincoln*

This study explores how public opinion regarding civil rights for Lesbian Gay Bisexual Transgender (LGBT) individuals has changed between 1992-2012. This 20-year period of attitudinal shift corresponded with the public debate, implementation, and eventual repeal of the restrictive Don't Ask, Don't Tell (DADT) personnel policy in the military. Using data from the General Social Survey (GSS) and American National Election Study (ANES), this paper examines the differences in attitudes towards LGBT civil rights between two employment sectors (military vs. civilian) and across several occupational groups. These differences in attitude are further examined across a 20-year time period. This paper seeks to provide historical context to changing attitudes beyond changes in demographic composition by building on previous research that explored the 'age-stability' and 'working class authoritarian' hypotheses to

explain trends in social attitudes. Attitudinal support is measured with a mean rating scale based upon latent variable and factor analysis. Preliminary analyses show that significant differences exist between employees in military and civilian sectors, and also between certain types of occupations. Sectoral and occupational differences in public opinion vary by specific aspects of the civil rights agenda, including free speech, military service, job discrimination, adoption, and gay marriage. These findings suggest key periods of attitudinal lag and lead associated with the public discourse of DADT, employment discrimination protections, and marriage equality legal battles. This also suggests an additive effect on the growing public support of LGBT civil rights, although the trends were not observed uniformly across occupations.

Sex in Surveys: How the Question Changes the Answer
Alian Kasabian, *University of Nebraska-Lincoln*

Researchers ask about sex categories (male and female) on almost every survey, and assume it is a static characteristic. What if the way we asked it changed the results? In this research, 457 people who responded to the 2014 Work, Life, and Gender web-survey were asked about sex and gender in several ways, including sex assigned at birth (75% female), are you: male/female (71% female), and which of the following best describes your gender identity (check all that apply – six options) – collapsed into man only (23%), woman only (67%), and any other combination (10%). Approximately 13% of the respondents produced answers that changed substantially across these three questions. When these answers are used as predictors of outcomes like job satisfaction, experiencing verbal harassment, and self-rated warmth, the outcomes for participants outside the binary were different from the others, and frequently in less positive directions (even when the majority of the group were highly educated and in professional jobs). Although sex is assigned at birth, and sex differences are well documented in many aspects of social life, the measurement of male/female does not always align with adult lived experiences, as evidenced by the dissimilarity of the non-binary group from the others in this research. It is common for research to use biological measures (sex) as proxies for social measures (gender), but this research demonstrates how the current standards can be both inconsistent (depending on wording) and inaccurate (as predictors of social outcomes).

May 15, 2016
Concurrent Session L

Assessing Survey Error for Mixed Mode Surveys

Seasonal Variation and Nonresponse Bias in a Population Survey of Health and Well-being
Jeanette Y. Ziegenfuss, *HealthPartners Institute*

In a monthly survey of 1,450 unique adult health plan members, we examine patterns of nonresponse by age, gender, race and health insurance plan type as well as survey month. In July 2016, HealthPartners began an ongoing survey of health and well-being of a random sample of their adult members. Selected sample are mailed up to two paper surveys followed by up to 6 telephone calls at various times of the day and days of the week in an attempt to complete a 10-15 minute survey. The paper survey is seven pages long and has 39 questions. Overall, the response rate is 25%. Differential nonresponse by frame characteristics are as expected with women, older, white and privately insured

individuals being more likely to respond to the survey. Interestingly, the relationship between likelihood to respond by these characteristics differs considerably over time. Over 12,000 pieces of sample are analyzed and conclusions are made about the differential impact of seasonal variation on subpopulations of HealthPartners membership, suggesting that time-limited nonresponse bias analysis may be inadequate for making decisions about the representative nature of a survey strategy. Further, some seasonal variation in collected survey responses may be due to underlying differences in relative nonresponse bias.

Modeling Representativeness Across Panels
Carina Cornesse, *University of Mannheim*
Tobias Enderle, *GESIS – Leibniz Institute for the Social Sciences*
Annelies G. Blom, *University of Mannheim*

In the past decade, probability-based online panels of the general population have arisen in Europe and the US in an attempt to combine the best of two worlds: the high coverage and response rates of face-to-face surveys with the timeliness, low measurement error and cost-effectiveness of the online mode. Yet, these panels differ in their sampling and panel recruitment procedures, including the treatment of sampling units that do not have computer and/or Internet access. This raises questions about their relative success in obtaining representativeness. In Germany, two panels - the German Internet Panel (GIP) and GESIS Panel - recruited their samples with different approaches to fieldwork and coverage in 2012 and 2013. Their recruitment data thus offer an ideal opportunity for researching the effect of different fieldwork strategies on panel representativeness. We compare response sets across panels in order to assess whether the GIP and GESIS Panel differ in their response set composition. Additionally, we draw comparisons across recruitment steps and panel waves within the GIP and GESIS Panel to assess the development of the response sets over the course of the panel life cycles. In our analyses, we use the sample-based R-Indicator as well as the Fraction of Missing Information (FMI) as measures for the representativeness of response sets compared to gross samples including respondents as well as nonrespondents. We base both measures on several sources of auxiliary variables, such as sampling frame information, interviewer observations during fieldwork, and commercial micro-geographic area data. In addition to the general sample-based R-Indicator, we compute partial R-Indicators in order to assess which characteristics drive the change in panel response sets across panel waves. Our results provide valuable insights into the effect of different sampling, fieldwork, and panel maintenance strategies on the composition of response sets over the course of panel life cycles.

Within-household Selection in Mail Surveys: Explicit Questions are Better than Cover Letter Instructions
Kristen Olson, *University of Nebraska-Lincoln*
Jolene D. Smyth, *University of Nebraska-Lincoln*

Randomly selecting a single adult within a household is one of the biggest challenges facing mail surveys. Yet obtaining a probability sample of adults within households is critical to having a probability sample of the US adult population. Prior studies have shown that at least 30% of adults who complete a mail survey are incorrectly selected, with selection accuracy falling greatly in households with 2 or more adults (e.g., Battaglia et al. 2008; Olson, Stange, and Smyth 2014). Thus, methods to improve within-household selection accuracy are critically important. In this paper, we experimentally test three alternative placements of the within-household selection instructions in the National Health, Wellbeing and Perspectives study (sample n=6000; respondent n=1002): (1) a standard cover letter informing the

household to ask the person with the next birthday to complete the survey (control); (2) the control cover letter plus an instruction on the front cover of the questionnaire itself to have the adult with the next birthday complete the survey; and (3) the control cover letter plus an explicit yes/no question asking whether the individual is the adult in the household who will have the next birthday. We find no significant differences in response rates across the three versions (letter only: AAPORRR1=17.8%; instruction: 16.5%, question: 15.8%). The explicit question, however, significantly improves selection accuracy relative to the other two designs (question: 66.7% accurate vs. letter only + instruction: 58.1% accurate). This paper will examine sample compositional differences across the treatments, differences between the sample composition and national benchmarks for each treatment, and whether theoretically-driven predictors are related to accuracy. The cover letter factor was also fully crossed with an incentive timing factor; as such, we will examine whether incentives moderate the effects of the within-household selection treatments. We will conclude with recommendations for survey practice.

Multi-mode Data Collection

Explicit vs. Implicit Data: Comparing Responses from a Web Survey to Behavioral Data Collected Directly from Smartphones

Noble Kuriakose, *SurveyMonkey*
Abhinav Agrawal, *SurveyMonkey*
Vincent Yiu, *SurveyMonkey*

A significant opportunity and challenge presented by big data is capitalizing on implicit information that's collected passively. Using SurveyMonkey Contribute, a panel of mobile device users who have agreed to allow data collection from their mobile phones, we conduct an experiment to understand how explicit data collected via a survey differ from implicit data gathered directly from respondents' mobile phones. We examine non-response bias, how recollection and memory differences affect behavior reports, and how demographic characteristics mediate these differences. We also consider the "fit for purpose" of surveys to gather data about how Americans use their mobile devices-- are some measures better (i.e. usage of apps, installing apps) to gather via a survey than others? Are reports of some apps or website more memorable and more reliable than others?

Effects of a Sequential Mixed Mode Panel Design on Nonresponse Error, Measurement Error and Total Survey Error

Mark Trappmann, *Institute for Employment Research (IAB)*
Joseph W. Sakshaug, *University of Manchester*
Manfred Antoni, *University Mannheim*
Reinhard Sauckel, *University Mannheim*

There has recently been an increase in the use of multiple modes of data collection within the same survey. Especially within panel surveys mixing modes has become a common form of data collection (e.g. HRS, UKHLS or GSOEP). Although mixed mode designs can be effective in reducing non-response bias or decreasing survey costs, a major problem with all mixed-mode designs is that it is difficult to separate (self-)selection into mode from measurement effects of the mode. These two error sources are always confounded. While techniques for separating self-selection effects from measurement effects have been proposed recently, they face the problem that in order to model selection into modes they can make use of those variables that are not themselves subject to mode specific measurement error. This leaves only external data (e.g. from the sampling frame, paradata or re-interviews) or variables for

which there is little evidence for mode specific measurement. Good validation data that allow a separation of These error sources are extremely rare, because the validation data have to be available for respondents as well as nonrespondents. In our presentation we will be able to make use of such data. PASS is a Major German household using a sequential mixed-mode design of CAPI followed up by CATI. For the complete sample we were permitted to link the survey's paradata (including final disposition codes and mode) to high quality administrative data on employment histories. In our presentation, we will first show how total bias develops before and after the mode change in the initial wave for several target variables. Next, we will disentangle this into a nonresponse bias and a measurement bias component. Finally we will show how the error sources are affected in later waves across the course of the panel.

An Examination of How Survey Mode Affect Eligibility, Response and Health Condition Reporting Rates in Household Surveys and Whether Length of the Questionnaire Mitigates These Results?

Michael J. Stern, *NORC at the University of Chicago*
Reem Ghandour, *Health Resources & Services Administration*

Due to decreasing rates of response for telephone surveys, a number of large, national level surveys have investigated or begun the transition to self-administered survey modes such as web and mail using address-based samples (ABS). However, the bulk of the research on mode effects is lacking in three ways. First, there are many studies that fail to use a true experimental design and instead compare results across a mixed mode survey to examine effects. Second, when there is an adequate design, the surveys tend to focus on cross-sectional local or student surveys, not national level results. Third, even when the two above requirements for accurate assessment are met, the effect of a screener for eligibility, sensitive questions, and the effects of questionnaire length are absent. To address this gap in the literature, we conducted a national level experiment using the National Survey of Children's Health (NSCH). The NSCH project involved a redesign of the historically telephone survey that incorporates two significant design changes over prior administrations: it moved from a random-digit-dial (RDD) to ABS frame and it moved from interviewer-administered telephone data collection to self-administered mail and web data collection. In an effort to anticipate changes in estimates due to mode effects, a mode effects experiment was designed and conducted wherein sampled households were screened and interviewed for the NSCH either on the Web, through the mail, or on the phone. In addition, within each modes respondents were randomly assigned an abbreviated or full length version of the survey. We compared eligibility, response rates, and health conditions reported by mode and considered the effects of questionnaire length in the models. The results shed light on sample variation, household size, reported health conditions, and eligibility all by mode and by questionnaire length providing researchers with considerations for the survey design.

What's Mode Got to Do With It? A Look at Different Interview Methods when Asking the Really Tough Questions

Kriston Koepp, *NORC at the University of Chicago*
Amanda Abramson, *NORC at the University of Chicago*
Shelly Plummer, *NORC at the University of Chicago*

NORC at the University of Chicago is conducting its third wave of the National Social Life, Health and Aging Project (NSHAP), a longitudinal, multi-modal study of approximately 4900 older adults born between 1920 and 1965 which includes an in-person interview, physical measure collection, and

respondent-administered leave-behind measures. As NSHAP is examining the interaction between aging, social relationships, and health outcomes, the interview consists of a constellation of topics such as sexual relationships, relationships with friends and family, religion, income, sexual interest, and the respondent's own thoughts and feelings. Due to the sensitive nature of these questions and interview time constraints, there are many interview modes that have been used over the past ten years of NSHAP to capture respondent's answers, such as computer-assisted personal interviewing (CAPI), computer-assisted self-interviewing (CASI), leave behind questionnaires (LBQ), and self-administered questionnaires (SAQ). This paper will take an in-depth look at which mode yields the greatest cooperation rates when dealing with sensitive issues and present results from NSHAP data collected between 2005 and 2015. In addition, this paper will look at respondent age, race, and gender and their corresponding effect on respondent cooperation rates between the four interview modes (CAPI, CASI, LBQ, and SAQ). Demographics from the respondent will be presented along with their correlation to response rate among modalities. This paper will discuss not only which interview mode or modes produces the highest success rate in an interview with sensitive topics but also whether demographics of respondents have a role in respondent cooperation rates across different interview modes.

Mixing Modes: Challenges (and Tradeoffs) of Adapting a Mailed Paper Survey to the Web

Sidney Wilkinson-Flicker, *American Institutes for Research*
Cameron B. McPhee, *American Institutes for Research*
Rebecca Medway, *American Institutes for Research*
Ashley Kaiser, *American Institutes for Research*
Katelyn Cutts, *American Institutes for Research*

Recent trends show that survey respondents are increasingly difficult and expensive to reach. Methodology research consistently demonstrates that tailored and adaptive designs may offer the best solution for collecting high-quality data. One strategy that can increase coverage and representativeness — and potentially reduce cost — uses sequential mixed-mode designs that include a web-based response component. In January 2016, the National Center for Education Statistics will test a sequential mixed-mode, web-push design for the 2016 administration of the National Household Education Survey (NHES). For several cycles, the NHES has used an address-based sample to administer a two-phase, self-administered mailed questionnaire in which sampled households are rostered using a phase-1 screener and then a single individual is sampled from responding households to complete a longer phase-2 "topical" survey. This presentation will describe the process of adapting the two-phase paper design to incorporate a variable-phase web survey, and some of the key challenges faced while transitioning from a well-tested paper-only to a mixed-mode administration. Authors will describe the tradeoffs between maintaining consistency with the paper instrument and optimizing the web survey; the complexity of building a web instrument that in some situations (e.g., single-adult households) must be a single-phase survey with both phases completed by one individual, while other situations require a different respondent to complete each phase; and the intricacies of using phase-1 screener data to customize wording in both English and Spanish using known information about the respondent. In addition to discussing the above challenges and proposed solutions, the paper will present selected results of usability testing and the resulting design changes to the web instrument. This study contributes to the growing body of research examining the most effective ways to use mixed-mode designs to increase survey response and representativeness while minimizing cost and mode effects in a national household survey.

Eye-tracking in Survey Research

Panel organizer: Aaron Maitland, *Westat*

Eye-tracking provides a window into how respondents process survey questions. As the name implies, eye-tracking tracks respondents' eye movement while they are reading and answering survey questions. It has been gaining popularity among survey researchers and has been used in the survey context for many purposes. For instance, eye-tracking has been used to evaluate and test survey questions as well as a survey's visual design. Eye-tracking has also been used to address theoretical questions about response behaviors and the underlying survey response process. One example is the use of eye-tracking to investigate the mechanisms underlying response order effects. Another example of the use of eye-tracking is to study visual context effects—how pictures and other non-textual elements on a computer screen affect survey answers. This panel showcases four applications of eye-tracking in survey context to address both practical and theoretical issues. The first paper deals with a practical question—the use of grid questions with 4th graders in the National Assessment of Education Progress. The other three papers address theoretical issues. The first of these papers explores whether eye-tracking can be used to measure response burden in a more objective manner. The next paper examines how respondents process pictorial examples differently from verbal examples and whether this difference affects survey answers. The final paper examines respondents' use of rating scales and explores mechanisms accounting for scale direction effects.

Evaluating Grid Questions for 4th Graders
Aaron Maitland, *Westat*

Eye-tracking has been used to better understand the survey response process. For instance, eye-tracking has been used to identify questions that are difficult to comprehend, how to present long lists of response options, and to measure the length of fixation on definitions in Web surveys. Grid questions have been commonly used in Web surveys as well as in other types of surveys. The literature shows respondents took less time to answer questions when they were presented in a grid than when they were presented individually across separate pages or screens. The use of grid questions, however, may also be associated with several undesirable outcomes, including higher breakoff rates, higher missing data rates, and straightlining. Relatively little is known about how children answer grid questions. This paper demonstrates how eye-tracking was used to determine the feasibility of using such questions to measure the background characteristics of students in the National Assessment of Educational Progress (NAEP) questionnaire. Fourth grade students were answered both grid and discrete (single-item per screen) versions of questions on tablet computers while wearing real-world eye tracking glasses. This study addresses four research questions related to the use of eye-tracking to test survey questions. First, we examine whether grid items require more effort to answer than discrete items for fourth grade students. Second, we investigate how the processing of sub items changes within a grid. Third, we examine how the processing of questions change over time. In order to address these research questions, we examine difference in the mean number of fixations per word and the mean duration per word for grid and discrete questions. Overall, the study finds support for the use of grid questions with fourth grade students in the NAEP. Implications for the use of eye-tracking equipment to evaluate survey questions are also discussed.

Use of Eye-tracking to Measure Response Burden

Ting Yan, *Westat*
Douglas Williams, *Westat*

Concerns about the burden that surveys place on respondents have a long history in the survey field. A review of the existing literature shows that the term "burden" is defined loosely and that researchers measure response burden in many different ways. Some measure response burden through properties of surveys/tasks that are believe to impose response burden, such as the length of an interview. Some measure response burden through respondents' attitudes and beliefs toward surveys, such as interest in and perceived importance of the survey. Others measure response burden through respondent behaviors (e.g., willingness to be re-interviewed) or direct respondent measures (e.g., feelings of burden). All three types of measurement are based on self-reports to survey questions, which are subject to the usual sources of reporting error due to misunderstanding of the survey questions, partial retrieval of information, biased or inaccurate judgment strategy, problems in mapping to the given response options, and more or less deliberate misreporting. In this paper, we will examine the use of eye-tracking equipment to measure burden. In eye-tracking research, task-evoked pupillary responses have been shown to be a consistent index of cognitive load and difficulty. For instance, dilated pupils are found to be indicative of higher levels of cognitive load and difficulty, in essence burden. We will create three measures—mean pupil dilation, peak pupil dilation, and latency to peak—as indicators of burden. These alternative measurements of response burden are free from errors in self-reports and are potentially stronger indicators of burden. We will evaluate the feasibility of using task‑evoked pupillary responses to measure burden in the survey context by comparing the three indicators to respondents' self-reports about burden.

The Effects of Pictorial vs. Verbal Examples on Survey Responses

Hanyu Sun, *Westat*
Jonas Bertling, *Educational Testing Service*
Debby Almonte, *Educational Testing Service*

Web surveys make it earlier to present images to the respondents than other modes of data collection. A few studies have examined the use of pictorial examples in Web surveys and found that the characteristic of the exemplars (e.g., their frequency or typicality) has an impact on the responses that are collected (e.g., Couper, Tourangeau, and Kenyon, 2004; Tourangeau, Conrad, Couper, and Ye, 2014). Tourangeau et al. (2014) compared visual examples with pictorial examples and found that respondents tended to report more foods consumption when they got verbal examples than when they got pictorial examples. The finding suggests that the pictures may narrow the interpretation of the category of interest. However, the findings also suggested that respondents are more likely to attend to the pictorial examples than to verbal examples. However, no direct evidence of respondent attention was collected to support either argument. Using eye-tracking, the current study compared verbal examples with pictorial examples in a lab setting to examine whether respondents attend to pictures more than words, whether items with verbal examples require more effort to answer than those with pictorial examples, and how the processing of items change over time. To address these research questions, we will examine differences in mean number of fixations and the mean duration for items with pictorial examples and verbal examples. The number of fixation is related to the amount of information that a respondent is processing, while the duration of fixations is related to the amount of difficulty that the respondent is having (Ares et al., 2014). The same food consumption questions with examples used in Tourangeau et al. (2014) will be used in current study.

Respondent Processing of Rating Scales and the Scale Direction Effect
Andrew Caporaso, *Westat*

Holding constant other scale features, the direction in which a scale is presented has been found to affect the resulting survey answers; respondents are more likely to select a scale point closer to the start of the scale regardless of its direction, producing primacy effects (Yan, 2015). What remains understudied is the mechanism underlying this scale direction effect. Two common response processing models are offered as possible explanations for these effects: satisficing, and anchoring and adjusting. The satisficing model treats the impact of scale direction as a special case of response order effect and argues that satisficers sequentially process the rating scale and select the first option that seems reasonable. The anchoring-and-adjustment heuristic assumes that respondents start with an initial anchor (the beginning of a scale) and make adjustments to the anchor until a plausible point is reached. Since both notions predict a primacy effect, it is hard to know which notion offers a better account for scale direction effect. To learn more about what's behind scale direction effects, we will collect eye tracking data from respondents' as they respond to a web survey. As the eye movement data (e.g., fixation counts and fixation duration) show directly the amount of attention paid to question components, we will first characterize how respondents process a rating scale and how the processing differs respondent characteristics. Then we will explore which of the two notions account for the scale direction effect. This paper demonstrates how eye-tracking can be used to address theoretical issues related to respondents' use of rating scales.

Optimizing Your Web Survey

Have You Taken Your Survey Yet? Optimum Interval for Reminders in Web Panel Surveys
Kirti N. Kanitkar, *The Gallup Organization*
Diana Liu, *The Gallup Organization*

Web survey mode is likened most closely to traditional mail mode in terms of participant experience. The conventional wisdom for web-based surveys has also borrowed from reminder intervals for mail surveys – i.e. reminders should be sent at weekly intervals (Dillman, 2000), although this view is now evolving (Callegaro, Manfreda, & Vehovar, 2015). With web surveys increasing in popularity exponentially, it's imperative to explore different reminder intervals to make optimal use of the opportunities of speed afforded by web surveys, without sacrificing data quality. Internal Gallup research indicates that a majority of web respondents respond to surveys within 24-48 hours of getting reminder emails. It might then be appropriate to reduce the interval between reminders for efficiency of data collection, if there are no differences in response rates, respondent characteristics, and substantive answers on the surveys. The present study addressed these questions through a set of experiments investigating the impact of shorter reminder intervals on response rates, composition of respondents and estimates of key questions in each treatment group. The study used data from the Gallup Panel, one of the few probability based mixed mode Panels which maintains demographic profiles of all its members. Eligible panelists were assigned to one of 4 treatment groups based on number of reminders (2 vs. 3) and reminder intervals after initial survey (2 days vs. 3 days). The fifth group was a control group with 2 reminders each one week apart, a schedule typically employed for Panel studies. The findings will explore response rates as well as estimates on various key questions by treatment group and detailed demographic profiles. Non-respondent characteristics and implications for sampling and survey field period decisions will also be discussed.

Impact of Field Period Length and Contact Attempts on Representativeness for Web Survey

Nick Bertoni, *Abt SRBI*
Chintan Turakhia, *Abt SRBI*
Robert Magaw, *Abt SRBI*
Allison Ackermann, *Abt SRBI*

Researchers do not have as much control over field period for self-administered surveys as they do for telephone surveys. Progressive increases in the use of Web as a data collection mode for general population surveys have resulted in growing interest in how long a web survey should be in the field and how many contact attempts should be made to ensure representation of various population subgroups. In this presentation, we examine field duration and contact attempts required for reaching various subgroups in a web survey. We examine the key demographics of early and late responders, as well as key survey variable measurements at various points in time. We utilize data from the Pew Research Center's American Trends Panel (ATP) for this analysis. ATP is a nationally representative probability web-based panel comprising of approximately 7,000 panelists. The field period for each wave generally lasts about 3 weeks. Our analysis is based on aggregated data for the last three waves of data collection for ATP. These analyses will inform survey researchers on the impact of contact protocol and survey field period on representativeness for web surveys.

Effects of Question Order and Paging in Online Surveys

Jillesa Gebhardt, *SurveyMonkey*
Noble Kuriakose, *SurveyMonkey*

The effects of question order, particularly the effects of asking a more general question before specific questions and vice versa, have been thoroughly researched and documented in the context of phone and face-to-face surveys. However, with the pervasiveness of online surveys, a review of previously studied phenomena within this new context is warranted. For example, do the effects of question order hold when respondents are able to see all questions on a page at the same time? This paper investigates the effect of paging and the interaction with question order in an online survey context. We examine four conditions in various combinations of question order (general to specific, specific to general) and page displays (more pages versus fewer pages). Findings from this study will contribute to the growing body of knowledge around best practices for online surveys.

Improving Data Quality in a Web Survey of Youth and Teens

Valrie M. Horton, *Abt SRBI*
Rosaella Branson, *Boys & Girls Clubs of America*
Benjamin Phillips, *Abt SRBI*
Elizabeth Fowlkes, *Boys & Girls Clubs of America*

Increasing prevalence of web data collection mode has led to considerable interest in measuring data quality in web surveys, especially when surveying youth and teens. In this paper, we examine data quality in a web survey context, using the Boys & Girls Clubs of America National Youth Outcomes Initiative (NYOI) member survey. The NYOI member survey is an annual survey of youth and teen members (ages 9-18) that measures academic success, good character and citizenship, healthy lifestyles, and club experience. With more than 100,000 online participants, the NYOI web dataset allows for

exploration of the impacts of a number of factors on data quality. We utilize various respondent level data in this analysis. This data includes language (the survey was offered in English and Spanish), age, grade, gender, race, device used for survey completion (computer or tablet), use of an audio component, and an honesty question. In addition, we also utilize Paradata on number of survey items answered, accumulated time spent on the survey as a whole, time of day survey was taken, type of question, and operating system. Data flags were created, post data collection, to identify fast respondents and respondents who straight-line or exhibit other patterned responses. We examine correlation between self-reported honesty, speed of survey completion, and patterned responses. We also examine the relationships that exist between fast/patterned responses and age, sex, race/ethnicity, household income status, and club engagement. Our conclusions help shed light on factors associated with survey response in surveys of children and provide guidance to other practitioners.

2020 Census Methodology

Panel organizer: Robert Colosi, *U.S. Census Bureau*

The Census Bureau issued the 2020 Census Operational plan in October 2015. Design changes and innovation abound in the documentation. The substantial investment in research from 2012-2015 provided the evidence needed to prove in innovations and contributed to new survey methods, as we cross into the 21st century. Innovative methods include building and updating our frame "In-Office", tailoring the contact strategy, implementing dynamic mode switching, using existing data to reduce field costs, and researching questionnaire content changes. Recognizing that some operational design remains unknown, research planned in 2016-2018 will help finalize the full 2020 Census design.

2020 Census Operational Design
Robert Colosi, *U.S. Census Bureau*

As the culmination of four years of research and significant investment in researching new methodology, the Census Bureau provides real innovation in methodology once again, through the issuance of the 2020 Census Operational Plan. Our research has focused on new methods for building and improving our address frame concluding using aerial imagery and administrative records data to add, delete, and change our address inventory. We will motivate self-response by making it easier to respond anytime, anywhere utilizing innovative methods that do not require census identifiers; utilize digital advertising and social media targeting; and contact strategies that includes Short Message Service (SMS/texting). By utilizing administrative records and third-party data, we will remove vacant units and some occupied units from expensive field enumeration visits to invest resources more efficiently. Research has focused on making our field staff more efficient including optimizing field assignments (time of day and number of visits) and routing; more efficient management structure; automated visibility into field activity; and redesigned quality assurance methods. Innovation in the design of the 2020 Census will improve methods, embrace technology, and reduce costs by over $5 billion.

Innovations in Frame Development for the 2020 Census
Laura Ferriera, *U.S. Census Bureau*

Measurement and reduction of coverage error is a difficult problem for census and surveys. The Census Bureau is innovating the 2020 Census design for frame improvement. We are implementing three components in the Address Canvassing Operation: In-Office Address Canvassing; In-Field Address

Canvassing; and an ongoing frame improvement operation that estimates national coverage of the Master Address File (MAF), called the MAF Coverage Study. This paper will focus on the new aspects of the Address Canvassing Operation along with preliminary results that helped with stakeholder decision making and plans for future testing. The In-Office Address Canvassing begins with geographic clerks comparing aerial imagery to current housing unit counts to detect change within blocks. When change is detected, In-Office Address Canvassing updates the MAF using imagery and multiple data sources including addresses from the Delivery Sequence File from the U.S. Postal Service. While some blocks are straight forward and easily completed in the office, other blocks are difficult to update using only in-office methodology. By the end of In-Office Address Canvassing, all blocks nationwide will be classified as current, updated to be current, or identified as questionable and sent to the field for updating. In-Field Address Canvassing will then send listers out to update addresses in blocks where In-Office Address Canvassing has indicated a need for fieldwork. The MAF Coverage Study is a new annual sample of blocks sent for fieldwork. The goals of the study are to generate estimates of the current differential coverage of the MAF, evaluate the accuracy of the In-Office Address Canvassing procedures, and update the MAF. The MAF Coverage Study will help us determine the impacts these innovations have on both the cost of frame development and quality of the address frame. Frame creation using these methods could improve address frame for both censuses and surveys.

Optimizing Self-response for the 2020 Census
Michael Bentley, *U.S. Census Bureau*

The decennial densus seeks to maximize household self-response rates to reduce costs and minimize the number of addresses requiring in-person followup. However there are many challenging factors, including declining survey response rates over time, distrust in government, an increasingly diverse and mobile population, and rapidly changing use of technology. After much research and testing we plan to implement new ways to optimize self-response. The 2020 Census will include several key methods to address these challenges and encourage households to self-respond. First, we plan to implement a customized communications and partnerships campaign to motivate people. Second, a tailored contact strategy of mailed invitation materials will be used, including mailing a paper questionnaire in the first mailing for some geographic areas with relatively lower Internet usage. We will offer an online self-response option that allows responses with and without a unique Census ID, to make it easy to respond anytime and anyplace. Multiple languages will be offered. Finally, Census Questionnaire Assistance will provide telephone help and collect interviews over the phone, and will also include web chat. Mixing modes and tailoring methodology will be core principles for the 2020 Census self-response.

Using Administrative Records to Reduce the Workload for Nonresponse Followup
Scott Konicki, *U.S. Census Bureau*

The Census Bureau is researching how to use administrative record information from government and other sources in place of field visits during the Nonresponse Followup (NRFU) operation. This paper describes an approach for identifying vacant and occupied housing units to be enumerated using administrative records during the NRFU. While the approach allows flexibility in balancing cost and quality, we show results that were utilized in determining the 2020 Census Operational plan.

2020 Census Content Research
Jennifer Kim, *U.S. Census Bureau*

Research and testing is currently underway to maximize content development and design of questionnaires across data collection modes, in order to help generate high response rates for 2020 Census operations. The refinement and finalization of the census content and question wording are critical, as 2020 Census topics must be submitted to Congress by early 2017, with the final question wording due April of 2018. Content development must also account for the demographic and cultural make-up of the U.S. that continues to increase in complexity, resulting in a growing number of households and individuals of Limited English Proficiency. Current research includes testing question alternatives for race and ethnicity, relationship, and coverage. The use of multiple languages is an important part of creating a census climate that will facilitate cooperation among census partners and the public at large; thereby increasing self-response, saving money, and increasing quality.

Current Affairs: U.S. and Abroad

Old Values, New Positions: Republican Reframing of Same-sex Marriage and Its Effects on Within-party Attitudes Pre and Post Obergefell vs. Hodges
Ashley A. Koning, *Rutgers University*

In June 2015, the U.S. Supreme Court's ruling in Obergefell v. Hodges effectively made same-sex marriage legal nationwide. The Court's decision was a major milestone for the issue, coming on the heels of an increasingly supportive public and backing from a rising number of elites in recent years. Most Americans are now in favor of same-sex marriage, with the exception of Republicans, who continue to struggle post-decision. A majority of Republicans at the mass level still oppose it, as do most Republican elites, with some members now encouraging the party to ignore the issue altogether. But other Republicans have been outright challenging the GOP's position on same-sex marriage legalization for years, acting as "atypical issue advocates." These atypical Republican advocates moreover express their support by framing their messages not in the same context as their left-leaning counterparts but rather around core conservative values that their own party already holds dear. These value-laden advocacy frames create an environment of "conflicting cues" where value orientation and party issue position seemingly do not match. I investigate whether these "conflicting cue" value frames are effective and have the ability to move opinion within groups where opinion was once thought immovable. I specifically analyze the "conflicting cue" value frames being used by Republican pro same-sex marriage advocates through two online surveys of self-identified Republicans nationwide. The surveys had identical value framing experiments but two different samples – one conducted immediately before and the other immediately after Obergefell v. Hodges. In each survey, Republican respondents were assigned to different pro same-sex marriage frames that manipulated both the featured partisan value and speaker's partisanship. Results indicate potential attitudinal shifts on the issue among Republicans both within and between the two surveys surrounding the decision. This work has theoretical implications for framing, opinion formation, partisanship, and values.

The Structure of Foreign Policy Attitudes in Comparative Perspective: Evidence from Four Countries

Timothy B. Gravelle, *University of Essex*
Jason A. Reifler, *University of Exeter*
Thomas J. Scotto, *University of Essex*

A recurring debate in the field of public opinion toward foreign policy relates to the number, content, and cross-national equivalence of the overarching concepts that serve to organize mass public attitudes toward foreign affairs. Put more succinctly, how many factors are there, what should we call them, and are they the same or different in different countries? This paper brings new survey data to bear on these questions. New surveys conducted in the United States, United Kingdom, France, and Germany provide an expanded set of foreign policy-related survey items that are analyzed using exploratory factor analysis (EFA) and exploratory structural equation modeling (ESEM). The paper argues for a more multi-faceted (and multi-factor) view of how mass publics structure their attitudes toward foreign affairs.

The DAV Veterans Pulse Survey: An Exploration of Cross-generational Differences in Veteran Attitudes and Opinions

Robert J. Torongo, *GfK Custom Research*
Ryan Tully, *GfK Custom Research*
Dan Clare, *DAV*

DAV (Disabled American Veterans), one of the nation's largest nonprofit, veterans service organizations, contracted with GfK to conduct a survey to better understand generational differences of America's veterans population. The DAV Veterans Pulse Survey provides the most complete picture to date of the attitudes and perceptions of veterans of all generations—from how they look back on their time in the military, to the challenges faced when they transition to civilian life, to their views on the benefits they receive, to how they feel the government and American public treat veterans. GfK's KnowledgePanel® is the most widely known and reputable probability-based web panel designed to be representative of the United States. This Internet-based survey was conducted using sample from GfK's KnowledgePanel; the survey was completed by 1,701 veterans. In this paper we explore and examine key response categories across service-era groups to understand diverse issues from veteran perceptions of service and its impact on their lives to the benefits they receive from the federal government. We will also explore challenges from the field in surveying veterans of different generations. The research contains both methodological and substantive implications for future studies research governmental program evaluation, health policy and outcomes, and outreach efforts for American veterans.

Personality, Perceptions of Democracy and Support for Authoritarian Alternatives in Mexico

Sergio C. Wals, *University of Nebraska-Lincoln*
Alejandro Moreno, *Instituto Tecnologico Autonomo de Mexico*
Jeffery J. Mondak, *University of Illinois Urbana*
Damarys J. Canache, *University of Illinois Urbana*
Dona-Gene Barton, *University of Nebraska-Lincoln*

Utilizing data from the Comparative National Elections Project (Mexico) 2012 and the Big Five of Personality framework, we assess the extent to which individual traits account for variance in perceptions of democracy. Specifically, we are interested in the relationship between traits and citizens'

evaluations of procedural efficiency of the political process and fairness during the most recent presidential election in Mexico. In addition, we explore the relationship between personality traits and individual preferences for non-conventional modes of political participation, including obstructionism and political violence. Finally, our study aims to shed light on the relationship between personality traits and individual preferences for authoritarian alternatives of government.

Documented or Undocumented? Qualitative Misperceptions about Immigrants in the United States
Daniel E. Herda, *Merrimack College*

Donald Trump's controversial statements about immigrants from Mexico have done little to hurt his Presidential campaign. His appeal is curious as those claims are demonstrably false. But, it raises questions about how many ordinary Americans hold similarly faulty perceptions about immigrants. The current study contributes to the growing literature on misperceptions about foreign-born and minority populations by examining whether respondents view the typical immigrant as documented or undocumented. I consider this qualitative perception alongside the more often studied quantitative perception of immigrant population size. Using data from the 2011 Transatlantic Trends Survey, I consider the extent, correlates, and potential consequences of holding incorrect perceptions. I find that a clear majority of Americans mistakenly believes that the typical immigrant is undocumented. Those holding this distorted view tend to be younger, suburban, and express greater dislike and threat from immigrants. Further, these warped perceptions about immigrant status are associated with greater opposition to birthright citizenship and provisions of the Dream Act. Interestingly, population size innumeracy, while common, yields much weaker effects. Overall, the findings suggest that much of the American populace is receptive to incorrect information as it fits within an already distorted and negative view of immigrants.

Cell Phones: Flags and States

The Effects of Sending Advance Letters to Cell Phone Respondents Using Appended Addresses
Kyley McGeeney, *Pew Research Center*
Courtney Kennedy, *Pew Research Center*
Jonathan Best, *Princeton Survey Research Associates International*
Missy Mosher, *Survey Sampling International*

One of the key challenges with cell phone random-digit-dial surveys has been the "cold call" nature of the recruitment. The researcher has little if any opportunity to explain the purpose and legitimacy of the survey before sampled adults start deciding whether they will answer the calls, much less participate. Thanks to a new advancement in cell RDD sampling, this dynamic is changing. Telephone sample vendors now offer the ability to sample on or append billing address and demographics to the cell phone sample. Researchers can use this information to send advance letters and incentives to future cell phone respondents. While promising, this possibility raises a number of research questions. What percent of the cell phone sample has a matched address and hat percent has matched demographics? What would be the effect of using these addresses and demographics to send a targeted advance mailing to cell phone sample addresses in terms of response rate, resulting sample composition and cost? To answer these questions the Pew Research Center conducted an experiment using a monthly political survey. Addresses and demographics were matched to the entire cell phone sample. Cell phone

numbers with a matched addresses and demographics were randomized to a treatment or a control group. The treatment group was sent an advance mailing containing a small prepaid cash incentive and a letter explaining that they would soon receive a phone call for a survey. The match rate of addresses and of demographics was examined Researchers also analyzed the response rate of the treatment and control groups, the demographics composition of the resulting samples and the cost associated with this design.

Across Space and Time: Diving Deeper into Cell Phone Activity Codes

Vanessa Maturo, *The Gallup Organization*
Stephanie Marken, *The Gallup Organization*
Missy Mosher, *Survey Sampling International*

As the percentage of cellphone only households continues to increase, researchers are forced to increase the percentage of cellphone completes in U.S. telephone studies, however cellphone completes are typically more expensive than landline completes because of manual dialing requirements and lower response rates among cellphone users. Researchers are turning to sample providers who offer screening services to help reduce total survey costs by removing nonworking cellphone numbers prior to delivery. For example, Survey Sampling International (SSI) appends activity codes to cellphone records to help researchers identify more productive phone numbers, categorizing cellphone numbers as: 1.) highly active (repeated outgoing activity over 1-12 months) 2.) previously active (activity tracked but not within 1-10 months) and 3.) inactive (no activity tracked for over 10 months or no activity ever identified). Previous research has explored flagging accuracy, productivity gains and the bias this screen introduces (Mosher and Best, 2014). This paper builds upon this research in two distinct areas. First, we examine screener accuracy by state, providing researchers important information about how they should adjust their approach based on their study population. Next, we examine the three activity codes to determine if their current categorization is ideal, and the extent to which researchers can gain efficiency while maintaining coverage by making further adjustments, such as categorizing records in the previously active group with activity tracked in 8-10 months ago in the inactive group. Finally, we estimate total savings by employing this screener on an RDD dual-frame cellphone/landline study of U.S. adults.

Age-targeted Flags for Use in Stratifying RDD Cell-phone Samples

Xian Tao, *NORC at the University of Chicago*
Benjamin Skalland, *NORC at the University of Chicago*
Vicki Pineau, *NORC at the University of Chicago*
David Yankey, *National Center for Immunization and Respiratory Diseases*
Zhen Zhao, *National Center for Immunization and Respiratory Diseases*

With more than two in every five American homes (45.4%) having only wireless telephones, it is important to include sample from a cell-phone frame in RDD telephone surveys to ensure sufficient coverage of the population of U.S. households and residents. However, cell-phone data collection can be expensive, particularly if targeting hard to reach sub-populations, e.g., households with children. The purpose of this paper is to assess the feasibility of using age-targeted flags in RDD cell-phone samples to improve sampling efficiency by stratifying the RDD cell-phone frame into an age-listed stratum with a higher expected concentration of age-eligible sample units and a balance stratum with a lower expected concentration of age-eligible sample units and oversampling cell-phone numbers from the age-listed stratum. A total of 10,000 RDD cell-phone lines fielded for the 2015 National Immunization Survey (NIS) were matched to four vendor's cell-phone age-listed databases to obtain flags indicating whether the

telephone number was expected to belong to the parent/guardian of a child less than 18 years old and, if so, expected to belong to the parent/guardian of a child less than 5 years old. For each flag, we used results from the NIS data collection operation to estimate the flag's coverage of the target population and the differential data collection costs between the age-listed and unlisted strata. These estimates are then used to produce an optimal sample allocation to the age-listed and unlisted strata. We estimated the cost savings that could be achieved through the use of the flags by comparing the cost per effective complete under the optimal allocation to the cost per effective complete when the flags are not used. Cost savings were estimated for each individual vendor's age-targeted flag and for combinations of the different vendors' flags.

Covering Our Most Mobile Users: Identifying Which States are Most Susceptible to Coverage Error and How Sampling Rate Centers in Adjoining States Can Help

Manas Chattopadhyay, *The Gallup Organization*
Stephanie Marken, *The Gallup Organization*

The percentage of cellphone only households and cellphone mostly users continues to increase nationally, and researchers are continually challenged to arrive at a sampling strategy that appropriate covers this increasingly mobile population. The percentage of cellphone users that reside in a different state than the state in which they purchased their cellphone poses a significant obstacle to properly sampling at the state level using an RDD cellphone frame. In their 2012 publication, Skalland et. al. demonstrated how this inaccuracy in the sample state and current state of residency varies tremendously by state for cellphone only users, and how undercoverage and overcoverage can impact final survey estimates. In 2015 Gallup built upon this research by sharing estimates of inaccuracy for all cellphone users by state, and information about how these differences can introduce bias into final survey estimates. In this paper, Gallup researchers provide updates of these estimates from a nationally representative survey of all U.S. adults. We also provide researchers with more information about which states with the highest undercoverage rates may benefit from sampling adjoining rate centers in neighboring states to minimize undercoverage error in final survey estimates. We also estimate total cost and productivity losses that result from sampling rate centers in adjoining states.

Movers and Shakers: Out-of-state Residents in Cell Phone Samples, BRFSS 2014

Pranesh P. Chowdhury, *U.S. Department of Health & Human Service*
Carol Pierannunzi, *Centers for Disease Control and Prevention*
Machell Town, *Centers for Disease Control and Prevention*
William S. Garvin, *Centers for Disease Control and Prevention*

BRFSS has been conducting telephone surveys using both landline and cellular telephone since 2011. Due to the portability of cellular telephone, many residents in a state retain cellular telephone phone numbers with area codes from other states. Since the BRFSS draws samples from all states, protocols dictate for the respondents to complete the core BRFSS interview and have the data transferred to the state of current residence. Therefore the number of persons who have moved into and out of each state (e.g., movers and shakers) while retaining their cellular telephone numbers can be tracked. The movers and shakers were compared with the respondents who had cellular telephone numbers sampled from their current state of residence by demographic factors, health care access, health behaviors, and chronic disease conditions. The percentage of out of state interviews ranged from 1.5% in Nebraska to 21.0% in Nevada (median: 5.8%). Compared with the respondents who had cellular telephone numbers that were sampled from their current state of residence, those with out-of-state interviews were

significantly more likely to be younger, disproportionately white non-Hispanics, college graduate, never married, and more likely to have health care coverage. Movers and shakers were less likely to report poor health, chronic conditions or risk behaviors (overweight or obese, smoking, lack of exercise) than their counterparts with cellular telephone numbers sampled from their current state of residence. Therefore, persons who might be left out of cellular telephone samples due to moving into and out of state may represent a potential for bias in cellular telephone surveys where transfer of data across state lines is not possible. Cellular telephone samples may include persons who have moved out of state, making them less efficient, but are not likely to include persons who have moved into states, resulting in under coverage.

Hot Button Issues in Public Opinion

Making It About Morals: Pope Francis Shifts the Climate Change Debate

Asheley R. Landrum, *University of Pennsylvania*
Robert B. Lull, *Annenberg Public Policy Center, University of Pennsylvania*
Heather Akin, *Annenberg Public Policy Center, University of Pennsylvania*
Kathleen Hall Jamieson, *Annenberg Public Policy Center, University of Pennsylvania*

Pope Francis's recent campaign emphasizing that acting to address climate change action is a moral obligation has received significant global attention. We examined whether the Pope was able to effectively leverage his credibility as preeminent moral leader to sway skeptics to consider climate change action from a moral perspective. In the current study, we conducted a national, cross-sectional survey of 1,441 respondents. Controlling for demographics religious and political beliefs, we analyzed the effects of climate change beliefs (e.g., the extent to which there is evidence of anthropogenic climate change), credibility ratings of Pope Francis, and the interaction between climate change beliefs and credibility on agreement with the Pope's message ("Humans have a moral obligation to act to address climate change"). As expected, climate change beliefs and credibility were positively related to agreement with the message. More importantly, there was an interaction between climate change beliefs and credibility: respondents with higher climate change beliefs agreed with the message, regardless of how credible they rated the Pope. However, among respondents with lower climate change beliefs, agreement with the message increased as credibility ratings increased. This suggests that the Pope's moral authority granted him some influence in shaping climate change beliefs from a moral perspective. In sum, there is evidence that Pope Francis was able to leverage his moral authority to reframe climate change action as a moral obligation. Whereas world leaders and climate scientists have tried to leverage political power and scientific evidence to shape climate change beliefs for years, Pope Francis's approach was different. His position as Pope grants him moral authority unique among world leaders, and from that authority he hoped to shift the climate change debate onto moral grounds. Our research suggests that he may have been successful.

Using Choice-based Conjoint Analysis to Inform Message Testing in an Environmental Context

Barry T. Radler, *University of Wisconsin-Madison*

Conjoint analysis is a multivariate technique that is used to understand how individuals develop preferences for complex multi-attribute stimuli. The objective of conjoint analysis is to determine what combination of a limited number of attributes is most influential on respondent choice or decision making. Unlike conventional estimation techniques that use direct scale ratings of individual and distinct

elements that are used to compose an attitudinal and decision model, conjoint analysis decomposes preference from global evaluations of whole stimuli bundles. This approach is uniquely predictive because it measures a common and realistic task; judging overall stimuli and making trade-offs among alternatives. The development of discrete choice models of conjoint analysis has further enhanced the mundane realism of the approach by asking individuals to choose instead of rate or rank. These advantages have made the technique widely used in marketing research to develop and evaluate products and product lines. Considering conjoint analysis has a proven track record of illuminating human decision making, its use in communication contexts has been limited. This presentation demonstrates a relatively novel application of discrete choice conjoint analysis to message testing in a social marketing context. Specifically, we review two projects which used conjoint to test the effectiveness of messages (posters) that encouraged conservation and environmentally responsible behavior. Using a realistic metric (choice), along with powerful Hierarchical-Bayes estimation and Latent Class Analysis segmentation techniques, these projects produced detailed models of message processing that distinguished among different market segments and allowed simulations to test different message combinations and competitive scenarios. Conjoint analysis provided greater confidence in the implementation of communications campaigns that can inspire and motivate action.

Measuring the Impact of the Great Recession in Generation X
Jon D. Miller, *University of Michigan*

Using data from the 28-year Longitudinal Study of American Life (LSAL), a measure of the impact of the Great Recession will be constructed. The respondents in the LSAL were 7th and 10th grade public school students in a stratified national probability sample and were 40 to 43 years of age in 2014 (N=4,200). The LSAL includes a comprehensive set of educational and occupational measures over the last 28 years and the 2014 annual survey included an extensive battery of items about the ways that respondents may have been impacted by the Great Recession (GR). A confirmatory factor analysis was used to construct a zero-to-10 scale of impact. A set of tables describe the factors associated with negative impacts from the GR and a structural equation model is used to estimate the relative influence of various social, educational, and contextual variables on the magnitude of the impact. In broad terms, the result demonstrate that individuals with higher levels of education and more years of occupational experience weathered the GR more successfully. There were some gender differences within lower educational strata but minimal gender differences among respondents with higher levels of education. The analysis also explores the impact of the GR on the ability of respondents to save or pay for the college education of their own children and to save for their senior years.