



A Meeting Place



2015 Conference Abstracts

May 14 - 17, 2015

The Diplomat • Hollywood, Florida

Thursday, May 14
1:30 p.m. – 3:00 p.m.
Concurrent Session A

Cross Cultural Measurement

Innovative Uses of Paradata Across Diverse Contexts

Beth-Ellen Pennell, *Institute for Social Research, University of Michigan*

Gina Cheung, *Institute for Social Research, University of Michigan*

This presentation will describe innovative uses of rich paradata to monitor production and quality indicators in household surveys in international settings. Data collection technologies that have been widely used in the United States for decades are increasingly being adapted and used in innovative ways in new contexts. This transfer of technology is facilitating new approaches that enable more efficient production and better quality control in these contexts through the collection and monitoring of rich paradata (process data). This diffusion of technology not only allows for immediate access to the survey and process data (including call records) to monitor field work quality but has also facilitated the use of other applications. These applications also include self-administered modes (e.g., audio computer-assisted self-interview [ACASI]), digital audio recordings, global positioning systems (GPS) for collecting contextual information or live tracking of interviewer travel to households, areal photography for sample selection, and the collection of various anthropometric data using digital devices, among other examples. With these innovations come new challenges, however. The presentation will discuss advantages, challenges and lessons learned across a diverse set of projects as well as make recommendations for new developments in this space. The presentation will provide examples from projects in five very different settings: China, Ghana, India, Nepal, and the Kingdom of Saudi Arabia.

Culturally-Related Response Styles for Attitude Questions: A Comparative Analysis of Chinese and American Respondents

Mengyang Wang, *University of Nebraska–Lincoln*

The existence of culturally-based survey response styles represents a major threat to the interpretation of research findings. Researchers conducting cross-cultural surveys are particularly challenged in drawing valid conclusions from survey data when respondents demonstrate culturally-based response patterns that are systematically different, though unrelated to content. In China, the 'Doctrine of the Mean' is a central aspect of Chinese culture. Unlike American culture, Chinese principles encourage people to express neutral, instead of polarized, attitudes. Although this cultural difference may have the potential to impact survey response patterns, there is little research studying underlying cultural effects by comparing Chinese and American response styles. In this study, a comparative analysis of responses from Chinese and American respondents are examined, using data from the 2013 Gallup World Poll, a multinational probability-based survey. To examine whether variable response styles for attitude questions can be explained by the presence of an underlying construct related to cultural difference, simultaneous latent class analysis is employed to examine the structure of perception patterns, as well as differences in patterns of Chinese and American respondents. Specifically, this research seeks to examine three issues: (1) whether Chinese respondents

AAPOR 70th Annual Conference

demonstrate significantly different response patterns compared to American respondents, (2) whether the occurrence of extreme response styles is lower among Chinese respondents, and (3) whether Chinese respondents are more likely to give non-opinion responses. Preliminary results indicate a complex pattern in which the likelihood American respondents opt for extreme responses are double the likelihood Chinese respondents do so across a range of topics. Moreover, Chinese respondents appear, on average, six times more likely than American respondents to select non-opinion response. These findings raise a number of issues that are important to address, in order to assess the problem of measurement artifacts unrelated to the constructs of interest before cross-cultural comparisons can be made.

Religion in the Soviet Bloc After the Fall of the Iron Curtain

Neli Esipova, *Gallup*

Dato Tsabutashvili, *Gallup*

This paper will examine people's attitudes toward religion in 29 former Soviet bloc countries, as well as their religious affiliation. While Soviet ideology sought to eliminate religion and replace it with universal atheism, the Catholic Church in Poland, for example, was able to resist. Given that religion was suppressed to different degrees in different countries under Communist rule, and is still suppressed in some of them today, religion is more important in some countries than others. People's tolerance of different religions also varies throughout the region. Using data from the Gallup World Poll collected annually between 2005 and 2014, we are able to shed more light on trends in religious affiliation as well as on the importance that people place on religion. Furthermore, we compare attitudes regarding religious tolerance. In all former Communist countries, the Gallup World Poll asks self-declared believers whether marrying someone outside their own religion is acceptable. In addition, we ask believers if they consider their religion to be the one true religion in the world or as just one among many equally valuable religions. Finally, the World Poll provides public opinion data on religious education in schools from across the region. And special attention is paid to attitudinal differences among various demographic groups, with a particular focus on age groups to compare those who grew up under Communist rule versus the younger generation.

Cognitive Testing of Survey Translations: Does Respondent Language Proficiency Matter?

Patricia L. Goerman, *U.S. Census Bureau*

Mikelyn Meyers, *U.S. Census Bureau*

Hyunjoo Park, *RTI International*

Alisu Schoua-Glusberg, *Research Support Services, Inc.*

Mandy Sha, *RTI International*

Historically, many researchers have followed the "rule of thumb" that translated survey instruments should be cognitively tested with only monolingual and limited source-language proficient respondents. There are various rationales for this practice, such as assumptions that 1) monolingual respondents are the intended users of translated survey instruments (Pan, et al., 2007); and 2) fully bilingual respondents might be more likely to understand a poorly translated phrase. Bilinguals are also more likely to understand inappropriately literal translations for concepts that do not exist in the target language. Whether pretesting translations with monolingual or bilingual respondents produces different results has rarely been studied empirically and most studies have not been designed to look at this issue specifically (Park, et al., 2014). This paper focuses on results from two different Spanish-language cognitive and usability studies with monolingual and bilingual Spanish-speaking respondents: 1) testing of a

AAPOR 70th Annual Conference

U.S. Census Test internet instrument and 2) testing of a U.S. Census Test CAPI instrument. In the first study, a small number of bilingual respondents were inadvertently included in the study. The second study was deliberately designed to compare testing with monolingual and bilingual respondents, with 20 interviews conducted with each type of respondent. Preliminary results from the first study indicate that bilingual respondents uncovered an equal or higher number of translation and usability problems as did the monolingual respondents. One limitation of these results is that the study was not specifically designed to examine this issue and thus bilingual respondents were not systematically recruited. In addition, there were correlations between English-language proficiency, computer proficiency and educational level, with the bilingual respondents having higher levels of each. We will compare the number and types of findings uncovered by monolingual and bilingual respondents in the internet and CAPI studies and will explore mode differences noted across the two studies.

New Direction in Sampling

Venue-Based and Real-Time Sampling Methodologies in an Intercept Survey of Cyclists

Olivia Saucier, *ICF International*
Ronaldo Iachan, *ICF International*
Heather Driscoll, *ICF International*

Intercept surveys can be useful for gathering information about travel to specific locations while the data is fresh in a respondent's mind (Schneider, 2013). They are valuable to transit agencies as a source of information on a cross-section of commuters. In fact, intercept designs may be the only cost-effective way to gather data about transit users or commuters where the incidence of these riders in the general population is low (Schaller, 2005). In partnership with the Bay Area Metropolitan Transportation Commission (MTC), ICF International designed and carried out an intercept survey of cyclists in nine Bay Area counties during the 2014 Bike to Work (BTW) Day. The survey was used to evaluate the impact of Bike to Work Day—specifically, its effectiveness at increasing bike ridership over the long term. This paper presents our sampling design and lessons learned from the field. The sample frame included more than 300 “Energizer Stations” located along commuting routes, where riders could stop for refreshments and free gear. Our multistage stratified sampling design selected a probability sample of stations and individuals (bike riders) that was representative of the BTW biker population in each county and area-wide. Trained interviewers were dispatched to 50 sampled Energizer Stations during the morning and afternoon commutes. They used a random selection method to approach riders and invite them to participate in a two-minute intercept survey, as well as a follow-up survey three months later. The intercept and follow-up waves achieved high levels of response, which was due to a mix of factors, including offering an incentive of high-value.

Sampling Wealthy Families in the Survey of Consumer Finances

Jesse Bricker, *Federal Reserve Board*
Alice Henriques, *Federal Reserve Board*
John Sabelhaus, *Federal Reserve Board*

As a survey of household wealth, the Survey of Consumer Finances (SCF) faces two main challenges. First, the concentrated distribution of wealth poses a challenge for any household

AAPOR 70th Annual Conference

wealth survey. The second, and more general challenge of all household surveys, is releasing timely data. To overcome the first challenge, the SCF has used data derived from administrative income records to over-sample wealthy families. Administrative data are used to predict household-level wealth, which is then the basis for sorting wealthy families into several high-end strata from which the over-sample is drawn. The over-sample is drawn from the timeliest administrative data. However, due to a variety of factors, using the timeliest frame data is leading the SCF field period to be completed much later than in the past. This paper first describes how the SCF uses administrative data to overcome the challenge of capturing the top-end families, and investigates two possible problems that can arise: failure to predict accurately model wealth from income, failure to achieve representative within-strata survey participation. In both cases, the SCF approach is successful: the models are accurate and respondents are representative. The analysis is based on comparing predicted wealth and income for respondent and non-respondent families within strata. This paper then describes how the SCF is proposing to overcome the second challenge: by using a sampling frame from the next-most recent year. Doing so will allow the SCF to start contacting over-sampled families earlier than in the past. However, expected family wealth is not necessarily stable across years. The second part of the paper analyzes the cost of this proposed frame change by examining the variability of wealth rankings across years.

Surveying the District of Columbia GLBT Community Using Respondent-Driven Sampling

Clyde Tucker, *American Institutes for Research*

Michael P. Cohen, *American Institutes for Research*

Angelina KewalRamani, *American Institutes for Research*

Sandy Eyster, *American Institutes for Research*

Respondent-driven sampling has been used to survey rare or hard-to-reach populations (Salganik & Heckathorn, 2004; Thompson, 1997). The District of Columbia (DC) Mayor's Office of GLBT Affairs, the DC Department of Health, and the American Institutes for Research (AIR) conducted a health survey of GLBT individuals in DC, a relatively rare and often hard to reach population. Respondent driven sampling was used to construct the sample. This paper discusses details of the sampling process, the issues that had to be addressed, the characteristics of the effective sample, and lessons learned. DC government agencies identified a number of GLBT organizations in DC; many of these organizations agreed to contact a sample of their members and request that they complete the survey. Each organization drew a sample of members using a sampling interval provided by AIR. These samples represented the first stage in the sampling process. The organizations were provided a letter that was sent to each sampled member explaining the survey's purpose, requesting participation, and giving a link to the web-based questionnaire. In addition to asking the respondents to complete the questionnaire, the letter asked them to provide the names and contact information of three friends who were part of the DC GLBT community but did not belong to their organization. These individuals comprised the second-stage sample. AIR sent letters similar to the ones sent in the first stage to these potential respondents requesting both their participation and three more names of GLBT residents of DC along with the contact information. This process continued until a sample of sufficient size was achieved. The goal of the process was to represent eligible individuals not belonging to participating organizations.

AAPOR 70th Annual Conference

Developing a Sampling Design for Exit Polls in States with Election Day In-Person, Early and By-Mail Voting

David B. Magleby, *Brigham Young University*

Geoff Cannon, *Brigham Young University*

Joseph Olsen, *Brigham Young University*

Daniel E. Williams, *Brigham Young University Statistics Department*

As states have expanded early and by-mail voting options, while often continuing to have election-day voting, the challenge of how to design a sample that accurately predicts the overall vote has become more complex. For more than three decades we have conducted a statewide exit poll designed largely to accurately predict the vote and provide data for academic analysis. In 2014 some more populous counties in our state opted to move to vote-by-mail and as a result we developed a sampling strategy that merged an election day sample design with a sample designed to learn about the voting behavior of early voters. One of the counties that shifted to a largely vote-by-mail system allowed voters to vote in-person on election day at voting centers. This county is part of two congressional districts which complicated data collection. In this county we experimented with I-pads as the means of data collection which allowed us to use zip codes to assign voters to the correct congressional district. This paper describes the methodology used in designing the sample and contacting the voters in a multimode survey approach. We contacted more than 20,000 people who had voted early by postcard asking them to go on-line and complete our survey. In addition we had 1,185 voters complete the survey over the phone. Non-response varied by mode of survey administration as did the sample size. Even with this added complexity the survey accurately predicted the closest congressional race in the state well within the margin of error. This paper describes what we learned in the process that may be of use to others who do exit polling of surveys using multimode designs.

Geo-Sampling: From Design to Implementation

Safaa R. Amer, *RTI International*

Designing a nationally representative probability based sample in developing countries can be challenging depending on availability of frames, resources, and survey timeline. Possibilities range from a large scale enumeration effort to construct a frame on one side and a random walk protocol on the other side. Our methodology represents an efficient and timely process which reduces the need for a full enumeration effort through the use a mix of Geographic Information System based information and census data. Geo-sampling also allows to avoid bias resulting from random walks. This paper outlines the steps of the methodology while highlighting its pros and cons as well as challenges faced in implementation as well as solutions developed while customizing the sampling design across different countries. Clustering and design effect are also explored and results compared to other traditional sampling methodologies.

Cell RDD Respondents Unmasked: Progress Report on Geo and Demo Appends to the Wireless Frame

Missy Mosher, *SSI*

Charles DiSogra, *Abt/SRBI*

Courtney Kennedy, *Abt/SRBI*

With the percentage of wireless-only households in the country topping 40% and wireless becoming the mode of contact preferred by the majority of the U.S. population, cell phone

AAPOR 70th Annual Conference

samples absolutely need to be included in dual frame samples in order to reach certain demographic segments. Unfortunately, with no source similar to landline white pages available, researchers using cell phone samples have had no choice but to use randomly generated lists with no accurate geographic or demographic information available for either selection or post-append. A number of compilers have begun building databases of cell phone numbers that include such information as name and address which then enables appending of low-level geographic and demographic selects and appends. We have seen studies in the recent past showing that the data included in these lists was very limited and potentially inaccurate. The expectation is that accuracy and coverage should improve over time, but has this happened? Our research evaluates the current status of appending low-level geography, like zip code and census tract and block group codes, to a randomly generated cell phone sample. Do current data sources improve our ability to predict a respondent's location? In addition, we examine the helpfulness of demographic appends like age and ethnicity for either targeting of a specific population for surveying or weighting results. Our hope is that the information available, including the availability of full addresses, has improved over time, and will continue to improve until research utilizing cell phone samples becomes much more cost efficient and effective.

Healthcare and the ACA: Enrollment, Transitions and Public Opinion

The Concentration of Health Care Expenditures in the U.S. and the Impact of Attitudes on Future Spending Levels

Steven B. Cohen, *Agency for Healthcare Research and Quality*

Estimates of health care expenses for the U.S. population are critical to policymakers and others concerned with access to medical care and the cost and sources of payment for that care. Medical care expenses, however, are highly concentrated among a relatively small proportion of individuals in the community population. Using information from the Medical Expenditure Panel Survey, this study provides detailed estimates of the concentration and persistence in the level of health care expenditures in the United States. Attention is given to identifying the characteristics of individuals with the highest levels of medical expenditures, in addition to those factors that are associated with low medical expense profiles. Analyses are included to discern the most salient factors that serve to predict the likelihood of experiencing high levels of medical expenditures in a subsequent year, in addition to the factors operational in predictions of experiencing low levels of medical expenditures in a subsequent year. Individual attitudes and opinions may visibly impact upon an individual's decisions on how and when to use health care services and associated decisions with respect to medical expenditures. These health care preferences also serve as important inputs in helping to predict health insurance coverage take-up decisions. Study findings continue to demonstrate that medical care expenses are highly concentrated among a relatively small proportion of individuals in the population. In 2012, the top 1% ranked by their health care expenses accounted for 22.7% of total health care expenditures with an annual mean expenditure of \$97,956 while the lower 50% accounted for only 2.7% of the total. Furthermore, the characteristics that distinguish these two disparate groups of individuals are notably distinct.

Opinions on Health Care Quality and Cost during ACA Implementation: Results from Three Surveys of the American Public and Employers

Jennifer Benz, *The Associated Press-NORC Center for Public Affairs Research*

Nicole E. Willcoxon, *The Associated Press-NORC Center for Public Affairs Research*

AAPOR 70th Annual Conference

Trevor Tompson, *The Associated Press-NORC Center for Public Affairs Research*
Emily Alvarez, *The Associated Press-NORC Center for Public Affairs Research*
Dan Malato, *The Associated Press-NORC Center for Public Affairs Research*
Becky Reimer, *The Associated Press-NORC Center for Public Affairs Research*
David Sterrett, *The Associated Press-NORC Center for Public Affairs Research*

The Associated Press-NORC Center for Public Affairs Research conducted three surveys in 2014 to gauge opinion among American adults, the privately insured age 18-64, and private sector employers during a crucial period of Affordable Care Act (ACA) implementation. This series of studies shows that for both consumers and employers, finding quality information about health care is a challenge, and cost continues to be a key factor for decision making. These studies, funded by The Robert Wood Johnson Foundation, provide a comprehensive picture of opinion on health care and insurance as the government makes major investments in making quality data more transparent and controlling health care costs, and as employers explore new benefits designs and adjust plan offerings to comply with the ACA. This paper will provide an overview of the opinion data collected across the three studies, and will focus on three key issues: 1) Health Care Quality: Less than a quarter of consumers are receiving provider quality information. Employers, too, are largely unfamiliar with objective quality metrics and instead tend to rely on data from health plans. 2) Health Care Costs: Health care costs are top-of-mind concerns for individuals and employers alike. Even among the privately insured, sizable proportions are worried about major unexpected expenses and avoid going to the doctor because of health costs. Employers are also worried about the bottom line, but they too place a high importance on employees' costs as they choose health insurance plans. 3) The Connection between Health Care Quality and Cost: About half of Americans believe that higher quality health care generally comes at a higher cost, while 37 percent say there is no real relationship between quality and cost. A majority of employers—though tentatively so—are willing to pay more for health plans that provide higher quality care.

The Affordable Care Act (ACA): Public Opinion Trends and Data Collection Challenges

Stephanie Marken Kafka, *Gallup*
Dan Witters, *Gallup*

In 2008, Gallup and Healthways initiated a partnership to track and understand the key factors that drive well-being. The Gallup-Healthways Well-Being Index 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, and a better understanding of the challenges of

AAPOR 70th Annual Conference

collecting and reporting these important data including question design decisions that affect the substantive data collected.

ZIP Code Tabulation Level Data: A New Way to Locate the Remaining Uninsured or Too Flawed to Be Useful?

Brett M. Fried, *University of Minnesota/SHADAC*

Michel Boudreaux, *University of Maryland*

Lynn Blewett, *University of Minnesota/SHADAC*

Elizabeth Lukanen, *University of Minnesota/SHADAC*

Karen Turner, *University of Minnesota/SHADAC*

In 2013, for the first time health insurance estimates became available for almost all ZIP Code tabulation areas (ZCTA) in the U.S. Although these events are not related, the release date of this data (the 2008-2012 American Community Survey (ACS)) coincided closely with the beginning of the first open enrollment period in the health insurance marketplaces. State officials at health insurance marketplaces immediately saw the potential for using data at the ZIP Code level to assist in enrollment efforts. After the first open enrollment period, this demand for ZIP Code level data has increased as health insurance marketplaces seek to improve their targeting of outreach and enrollment efforts to reach the remaining uninsured. In 2014, after the close of the first open enrollment period, Kaiser estimated that only 28% of the potential marketplace population had been enrolled. In this paper, we use a unique data set composed of plan selections at the ZIP Code level for 36 states that has been mapped to ZCTA level data from the Census. We use this data in a regression framework to identify what community level characteristics, including levels of pre-ACA uninsurance, race, poverty and other characteristics are associated with plan selection. The average number of plan selections per ZCTA is 436 and there are a total of 12,010 ZCTAs that have 50 or more plan selections. Preliminary results suggest that income, health insurance and race/ethnicity are associated with plan selection. We report results using interactive maps that include multiple levels of estimates and geographies. There are other organizations engaged in similar efforts but none to our knowledge that provide our level of statistical rigor, transparency and support. This research provides a blueprint for other researchers interested in how best to translate data at low levels of geography into useful policy relevant actionable information.

Promises and Perils of Re-Contact Surveys: Strategies for Studies of Low Incidence and Hard-to-Reach Populations

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

Susan Sherr, *SSRS*

Robyn Rapoport, *SSRS*

Lewis Raynor, *Oregon Health Authority, State of Oregon*

Rebekah Gould, *Oregon Health Authority, State of Oregon*

Kathleen Thiede Call, *State Health Access Data Assistance Center (SHADAC),
University of Minnesota*

Major policy changes, such as the Affordable Care Act (ACA), directly affect a relatively small portion of the population. Survey research provides a method of measuring early impacts; however reaching the relevant respondents is a difficult and expensive enterprise. One option to reduce expense and increase response is to recontact those with a known health insurance status prior to ACA implementation, and ask whether there have been changes in their coverage and why. To measure effects of the ACA, Minnesota and Oregon recontacted

AAPOR 70th Annual Conference

respondents previously identified as uninsured and/or individual insurance purchasers (the relevant populations for the ACA). People without health insurance are a low-incidence population (less than 10% in Minnesota and 14% in Oregon), and demographics of the uninsured population tend to mirror that of populations with lower response propensity. Fortunately, Minnesota and Oregon regularly conduct surveys of health insurance coverage and access providing a data source for contacting this hard-to-reach group. Minnesota uses a dual-frame random digit dial (RDD) telephone survey. Oregon uses an address-based sample (ABS) design and allows respondents to complete the survey in multiple modes. Goals of the recontact studies included determining: current health insurance coverage; pathways to insurance coverage; reasons for coverage or non-coverage; and barriers to gaining coverage. A recontact methodology has pros and cons for measuring change that can be attributable to ACA implementation. On the positive side, the sample is comprised of cooperative respondents, reducing costs associated with locating these people in the general population. In addition, households whose residents (e.g., those with lower incomes) are likely to have the health insurance status of greatest interest to a researcher can be pinpointed. The negatives include inconsistent information over time and higher nonresponse among members of the target groups of interest. This paper highlights the planning processes and reports on study outcomes.

Welcome to the Mainstream? Same-Sex Marriage, Sexual Orientation and Gender Identity

Changing Attitudes Toward Same-Sex Marriage in the United States – 1988 to 2012

Duane F. Alwin, *Penn State University*

Kyler J. Sherman-Wilkins, *Penn State University*

This paper examines changes in Americans' acceptance of same-sex marriage from the late 1980s through 2012. We discuss these trends against the backdrop of recent judicial and state-level legislative actions that increasingly permit marriages between persons of the same sex. To empirically examine these changes, we use data from five surveys from the General Social Surveys (GSS) over this period to show the dramatic change in the direction of increasing support for same-sex marriage. We locate these trends in two major sources of change—inter-cohort (generational) differences and intra-cohort (historical) differences in such attitudes over time. We re-examine the conclusions of prior research that the majority of the trends in attitudes toward same sex marriage have come about through 'period' or 'time-related' events, rather than through the replacement of older generations by younger ones that have more tolerant views (i.e. cohort replacement). In examining these historical patterns, we focus on two related explanatory arguments. First, we investigate the extent to which changes in tolerance toward homosexuality were a leading indicator of later shifts in attitudes to same-sex marriage. A substantial amount of literature has suggested that changes in the tolerance about homosexuality followed a cohort replacement pattern, and to the extent this is the case, attitudes toward same-sex marriage may have followed a similar pattern as a consequence of their dependence on a more global structure of tolerance toward homosexuals. Second, we examine Rosenfeld's demographic "age of independence" hypothesis, which holds that beginning with the cohorts coming of age in the 1960s, individuals experienced reduced parental control over their children's romantic relationships, resulting in greater tolerance of homosexuality and attitudes toward same-sex marriage, views that were more easily averted by previous generations of parents. Our analysis finds support for both hypotheses, and we discuss the implications of these results.

U.S. Hispanic Receptivity to Self-Reported Measure of Sexual Orientation

Christine Kudisch, *Experian Marketing Services*

Max Kilger, *Experian Marketing Services*

Josephine Leonard, *Experian Marketing Services*

Charles D. Palit, *University of Wisconsin*

In 2003, Experian Marketing Services conducted an independent test to examine mail survey return rates and item non-response effects when a portion of the sample was exposed to a self-reported measure of sexual orientation. The findings from this earlier study suggested that U.S. Hispanics may be more reluctant to return a survey with a question asking sexual orientation in it compared with non-Hispanics. More than a decade later, in light of changing attitudes, perceptions and other environmental conditions, we echo a similar test among U.S. Hispanics using Experian Marketing Services' Simmons National Consumer Study, a national survey of U.S. consumers reporting data on a national probability sample of approximately 25,000 total adults 18+ including 7,500 Hispanic adults age 18+ annually. Using an experimental design, a random selection of 10% of the Hispanic sample which includes both English-language and Spanish-language Hispanics is exposed to the sexual orientation measurement. This presentation will examine the effects on survey return rates and item non-response to the sexual orientation measurement among U.S. Hispanics in general but will also take a closer look at potential effects to these same metrics by other Hispanic characteristics which are often closely associated with levels of acculturation in the U.S., including but not limited to language use, country of origin and whether the respondent was born inside or outside the U.S.

Assessing Progress in the Measurement of Sexual Orientation and Gender Identity in Population Surveys

Stuart Michaels, *NORC at the University of Chicago*

Michael Stern, *NORC at the University of Chicago*

Measurement of sexual orientation and gender identity across the full range of the U.S. population pose important methodological challenges. At the same time, we have come a long way from the late 1980s when there were no representative samples that asked questions about sexual behavior and identity of the adult population in the U.S. In 1988 the Gender Social Survey (GSS) included a set of self-administered questions about sex of sexual partners for the first time and have included them in every round since. In 1992 the National Health and Social Life Survey (NHSLs) collected data on a national sample of 18-59 year olds including sexual behavior, identity and attraction. In 2002, the National Survey of Family Growth (NSFG), a large national survey of 15-44 year old men and women began including questions on sex of sexual partners, sexual identity, and sexual attraction. At the same time, the National Health & Nutrition Examination Survey (NHANES) included questions about sexual behavior and identity. Beginning in 2013, the National Health Interview Survey included a refined sexual identity question on the full adult population 18 and above. Gender Identity questions have not been asked in any national general population survey but questions have been asked of representative samples in two states, Massachusetts and Vermont. This paper assesses the current state of the art of measuring sexuality and gender in representative surveys by reviewing and comparing question and questionnaire design issues as well as the findings from major population surveys. It will also compare variation in estimates across surveys in relation to variation in questionnaire and mode as well as variation in non-response by population subgroups defined by factors such as sex, age, education, and race/ethnicity.

**What Sex Were You Assigned At Birth on Your Original Birth Certificate?
Methods of Identifying Transgender Respondents in a District of Columbia Health Survey**

Angelina N. KewalRamani, *American Institutes for Research*

Clyde Tucker, *American Institutes for Research*

Sandy Eyster, *American Institutes for Research*

Jeffrey Poirier, *American Institutes for Research*

Prior research has shown that the order in which questions are asked in a survey can have a significant impact on the results. Question order and context are especially important when considering questions related to gender identity, gender expression, and sexual orientation. This paper explores whether varying the order of a two-question gender identity series impacts the identification of transgender respondents. The District of Columbia (DC) Mayor's Office of GLBT Affairs, the DC Department of Health, and the American Institutes for Research (AIR) conducted a health survey of Gay, Lesbian, Bisexual, and Transgender (GLBT) adults living in DC. The study used respondent driven sampling (RDS) to build a sampling frame of GLBT individuals and included a web-based self-administered questionnaire. The health survey collected assigned sex at birth (male or female) and current gender identity (e.g. man, woman, transgender) in two separate questions. This is frequently referred to as the "two-step" approach since it uses two questions to classify respondents as transgender or cisgender (non-transgender). According to a 2012 study, the two-step approach was more successful in identifying transgender respondents than a single gender identity item that included a transgender response option (Tate et al., 2012). Despite the wide use of the two-step approach, it is unclear whether assigned sex at birth should be asked before or after current gender identity, as no research has been conducted on ordering effects of these two items (Sausa et al., 2009; Tate et al., 2012). In this paper, we examine the results from an experiment that employed a split-ballot design to test whether the order of the questions changed the proportion of respondents identifying as transgender. We conduct additional analyses to determine whether the order of the gender identity items influenced reporting on subsequent sexual orientation items.

Patterns of Response and Nonresponse to Sexual Orientation Measures

Jennifer M. Bouterse, *Experian Marketing Services*

Max Kilger, *Experian Marketing Services*

Josephine Leonard, *Experian Marketing Services*

Historically the measurement of sexual orientation has proved difficult for survey researchers. The sensitivity of the subject matter has traditionally been a significant barrier in obtaining responses to sexual orientation measures. In addition, the concept of sexual identification itself has been one that has housed significant debate among researchers in terms of understanding how individuals arrive at a personal assessment of their own sexual orientation. In this paper we explore the issue of item nonresponse to a measure of sexual orientation contained within a very large national probability survey. We begin by discussing historical difficulties in defining and measuring sexual orientation and then outline the strategy chosen to minimize nonresponse. We then proceed to provide some descriptive analyses of item nonresponse rates for some key demographic and lifestyle variables and conclude this section with a multivariate model of item nonresponse to this sensitive question. The next analysis examines the potential "nonresponse contagion effect" that a sensitive measure such as sexual orientation may have on measures that surround the measure in question. The hypothesis here is that respondents,

AAPOR 70th Annual Conference

especially in larger surveys, may skip measures in close proximity to the sensitive question when they also are nonresponders to the sensitive question itself. Our final effort is a more qualitative one where we describe the results of visually examining respondent responses to the sexual orientation measure. It has been our experience that given both the sensitivity and the difficulties in responding to this measure that respondents often respond with written comments that reflect a series of themes including providing response categories not listed as possible responses, remarks on the sensitivity of the subject, subjective remarks on the morality of the lifestyle among others.

HIV/AIDS in the Lives of Gay and Bisexual Men in the US: Factors Associated with Getting Tested

Mira Norton, *Kaiser Family Foundation*

Jamie Firth, *Kaiser Family Foundation*

Sarah Levine, *Kaiser Family Foundation*

Liz Hamel, *Kaiser Family Foundation*

Tina Hoff, *Kaiser Family Foundation*

Jennifer Kates, *Kaiser Family Foundation*

Mollyann Brodie, *Kaiser Family Foundation*

While gay men make up just 2 percent of the U.S. population, they account for two thirds of new HIV infections, a majority of people living with HIV, more than half of all AIDS deaths since the epidemic's beginning, and gay men are the only group in the country among whom new infections are on the rise. To better understand this population's knowledge of, attitudes about, and experiences with HIV/AIDS, the Kaiser Family Foundation conducted a survey of 431 men who self-identified as either gay or bisexual, using GfK's nationally representative, probability-based Internet panel. The GfK KnowledgePanel's annual profile includes questions about sexual orientation, which allowed for the sampling of these previously self-identified gay and bisexual men. Few representative surveys have been done with this population, none of which have focused specifically on HIV/AIDS. Relatively few gay and bisexual men report getting tested for HIV as regularly as advised and three in ten say they have never been tested. Even larger shares of those under age 35 and men of color report never getting tested. This paper will use logistic regression analysis to explore what factors are associated with HIV testing among gay and bisexual men, including opinion and attitudes about HIV/AIDS, the impact and experience of stigma, and connection to the medical system. Additionally, this paper will highlight potential avenues for outreach efforts to engage and educate gay and bisexual men about HIV testing.

Web: Metrics, Usability and Quality

Comparing Field and Laboratory Usability Tests to Assess the Consistency and Mistakes in Web Survey Navigation

Nicole Gonzalez, *NORC at the University of Chicago*

Michael Stern, *NORC at the University of Chicago*

Abigail Jones, *NORC at the University of Chicago*

Reem Ghandour, *Office of Epidemiology & Research, MCHB Health Resources & Services Administration*

Alyson Croen, *NORC at the University of Chicago*

The increase in internet penetration coupled with smartphone and tablet use makes the web surveys an attractive mode; however, there exist questionnaire design issues, as well as device-specific visual design concerns, that are poorly understood. Historically, researchers have assessed measurement error caused by survey mode effects through experimental field tests where several versions of a question are randomly assigned to a subsample of respondents with different modes. As more web and multimode surveys have been used in the general population, ways of uncovering potential areas of measurement error have been developed. One such technique is usability testing. Usability testing focuses on the human and computer interaction. Thus, it is not simply about the respondent's personal opinion or thoughts about the web survey or any given item, but instead focuses on objective results including, for instance, the time taken to answer a question, whether respondents change their answer, and whether respondents make mistakes in navigation or usage of the web survey. As Couper (2011, p. 385) explains, usability testing, or user-centered design "is more than user friendliness." In this paper, we assess the findings from a study where a complex telephone survey was redesigned for web and mail. We conducted 90 usability tests as well as fielding a large mode effects experiment using separate samples for web, mail, and telephone. Our research sheds light on the impact of visual complexity of rostering questions, matrices, and utilization of question grouping in web surveys, as well as the comparability of different visual groupings among multiple devices with different screen sizes, while also highlighting what issues are uncovered by usability tests, field tests, and both.

Evaluating Visual Design Elements for Data Collection and Panelist Engagement

Leah Christian, *The Nielsen Company*

Casey Langer Tesfaye, *The Nielsen Company*

Tom Wells, *The Nielsen Company*

Darin Harm, *The Nielsen Company*

The visual design and layout of websites and online data collection tools influences how people perceive, navigate and understand the information presented, as well as how they interact with the instrument. Visual design principles can be used to help the user recognize the most important information on the page and guide how they interact with specific features. For this study we use visual design principles to inform the design of a website that is intended to provide and collect information from research participants in a longitudinal panel and then evaluate the design of the website. The measures we use to evaluate the design include both passive measures, such as eye-tracking and performance data, and active measures, such as user think-aloud comments and answers to standardized questions in order to understand how people perceive the information on the site and navigate certain features. We will focus these findings on the evaluation of specific components on the site, including the organization of

AAPOR 70th Annual Conference

information, navigation tabs, use of images, card displays and edit buttons. Since this website will ultimately be used by a diverse population, the influence of respondent characteristics (e.g., age, race, computer literacy), and how to account for these characteristics through visual design, will also be discussed. We will present quantitative and qualitative findings related to each of these components and discuss how the information is used to improve the visual design of this website, as well as other websites with similar goals. In addition, the findings from the study will be used to provide best practices or guidelines for presenting particular types of information on websites and online data collection instruments.

The Use of Mobile Devices to Track Family Interactions

Faith M. Lewis, *Abt SRBI*

Ricki Jarmon, *Abt SRBI*

JoAnna Hunter, *MDRC*

Donna DeMarco, *Abt Associates*

Debi McInnis, *Abt Associates*

Ray Hildonen, *Abt SRBI*

Daily diary measurement studies are commonly used in the behavioral and social sciences to capture information on experiences as they happen. The Supporting Healthy Marriage Study was the first large-scale, longitudinal experimental evaluation of marriage education programs for low-income, racially and ethnically diverse married couples. Supported by the Administration for Children and Families within the U.S. Department of Health and Human Services, the evaluation was led by MDRC with Abt Associates, SRBI and other partners. One component of this study, the Supporting Healthy Marriage (SHM) Daily Diary Study, funded by the WT Grant Foundation, took the diary measurement approach further by using mobile technology to capture sensitive data about family interactions and disagreements from multiple members in a household – mother, father and child – over a 15 day period of time. The study captured data on the daily interactions of 100 families comprising 300 respondents across multiple sites. Each participant was given a handheld device for daily reporting. Participant training was conducted via telephone, and reminder calls were made to non-respondents. The benefits to a mobile diary approach are that it allows researchers to collect data in “real time,” improving recall and data quality. Mobile diaries are more portable and discreet for respondents, and real time collection allows researchers to access reporting as it occurs. However, collecting self-reported data from family members on sensitive data such as interpersonal communications and disagreements on a daily basis through mobile technology posed several recruitment, training, and technology challenges. It was especially challenging to obtain consent because of concerns about privacy and digital safety risks to children. This paper summarizes the challenges specific to this study and generalizable to diary studies as a whole and presents the innovative ways the research team overcame them.

Return To Sender: An Evaluation of Undeliverable (e)Mail in the Modern Age

Jenny Marlar, *Gallup*

Daniela Yu, *Gallup*

Internet surveys are becoming increasingly popular, yet response rates continue to remain low for the majority of studies. For many studies, response rates and participation rates may be impacted by addresses that are undeliverable. Emails may be returned to the sender for a variety of reasons - inaccurate address, full inbox, or technological failures. These bounce backs are likely to be missing completely at random. It is also common for email providers to block email or divert bulk messages to SPAM boxes. When a provider blocks all respondents

AAPOR 70th Annual Conference

from receiving the email, it is hypothesized that these respondents are not missing at random from the sample frame. The current study utilizes data from the Gallup Panel, a probability-based, multimode panel with over 50,000 members who respond to surveys via the internet. Email address and demographics are known for all panel members. The eight largest email domains in the panel, each representing at least 2% of all panel members, were analyzed to explore the demographic profile of each email domain. Significant demographic differences were found between domains, with some domains more likely to include younger, male, minority, and lower educated respondents. The study also explored the substantive results of five panel studies. Respondents from each domain were individually excluded from the results to create sets of hypothetical datasets that simulated an exclusion or block for each domain. Each study had eight sets of analysis, for a total of 40 simulated datasets, that covered a wide variety of topics. Despite the demographic differences, few questions had meaningfully different results when an entire domain was excluded.

Methodological Briefs: Advances in Designing Questions in Brief

Measurements of Adiposity: Methodologies, Potential Measurement Error and New Measurement Techniques

Heidi Guyer, *Survey Research Center, University of Michigan*

The rate of overweight and obese individuals continues to rise in the U.S. and other countries and the association with increased risk for multiple health factors is ever present. Additionally, weight is a common confounder that is controlled for in most analyses of health-related outcomes. With this in mind, accurate measurement of adiposity and weight are important. Surveys have often relied on self-reported height and weight, from which body mass index (BMI) is calculated and weight status categorized as below normal, normal, overweight or obese according to set criteria. Self-reports are inexpensive to collect and often provide reliable information with little missing data. However, measurement error can lead to misclassification of weight status and biased associations. This presentation will provide an overview of measurement techniques for measuring weight and adiposity in survey research taking into account potential measurement error, mode of administration and associated costs. Validation results from studies with both self-reported height and weight as well as actual measures of weight will be presented. Additionally, an overview of new opportunities that novel technical advances offer in the field of survey research will be discussed.

Asking Sensitive Questions in a GLBT Health Survey. Does Instruction Placement and Length Matter?

Mark Masterton, *American Institutes for Research*

Mengmeng Zhang, *American Institutes for Research*

Survey questions regarding physical and emotional abuse are notably problematic for survey researchers. These questions typically include a lengthy set of definitions and instructions below the question stem in order to measure the frequency of very specific types of events. However, past research indicates that long and complex instructions are more likely to be ignored introducing data quality issues. Other research suggests that embedding instructions within or before a question stem improves the likelihood that these instructions will be followed. This research has focused on broad topics that are applicable to the general population. There has been little comparable research to determine if these findings are consistent when examined for

AAPOR 70th Annual Conference

sensitive topics, such as physical and emotional abuse. This research utilizes data from a 2014 web-based self-administered sample survey of the GLBT population of Washington, DC to examine the impact of characteristics of instructions for items related to abuse on survey responses. This research includes two independent split panels focusing on items related to incidents of abuse. One split panel experiment focuses on items pertaining to domestic abuse and abuse due to sexual orientation. Half of the sample will receive a version of the item with instructions before the question stem, while the other half of the sample will receive a version of the item with instructions after the question stem. The other split panel experiment focuses on items related to sexual abuse and threatened sexual abuse. The sample will be split so that half of the respondents receive longer versions of the instructions and definitions and the other half receive abbreviated versions of the instructions and definitions. First, we compare responses between each treatment group, controlling for respondent characteristics. We then discuss strategies to minimize the impact of instruction context and question content on sensitive questions.

Who Really Thinks Global Warming is Happening: An Experiment on Response Order Effects in Attitude Towards Global Warming

Sujata Pal, *Abt SRBI*

Mindy Rhindress, *Abt SRBI*

Gemma Natori, *Abt SRBI*

Geoffrey Feinberg, *Yale University*

In the last few decades, several surveys have addressed public opinion and knowledge of global warming and climate change issues. These polls generally concur that Americans are increasingly more aware of the issues involving climate change although most see it as a distant rather than imminent threat (Nisbet and Myers, 2007). A review of these surveys also reveals the need for further methodological research that explores the role of question wording, format and data collection mode on survey results in this field. For example, one study found that framing the question as “global warming” as opposed to “climate change” has differential effect on public opinion based on political persuasion (Schuldt, Konrath, and Schwarz 2009). In the proposed paper, we explore whether reported response to the statement “global warming isn’t happening” is associated with the order in which the response options are presented. In survey literature, such response order effects have been shown to create a “primacy effect” or a “recency effect” (Krosnick and Alwin, 1987). We used data from a representative survey of American’s attitude towards global warming conducted by Abt SRBI for the Yale Project on Climate Change Communication (YPCCC). The survey was conducted by telephone using a dual frame landline and cell phone sample design. A split-ballot experiment was conducted in which respondents were randomly read the response options in two different orders. In one order, the statement “global warming isn’t happening” was presented first. In another order, the response options were reversed and this statement was read at the end. We explore whether this led to any “response order effect”. We also look into possible relationships with relevant demographic, political and attitudinal variables.

Varying Administration of a Sensitive Question to Reduce Item Nonresponse in a Telephone Survey

Julie Pacer, *Abt SRBI*

Martina Smith, *Abt SRBI*

Kelly Daley, *Abt SRBI*

Survey questions about sexual orientation are recent additions to federally-supported general population surveys such as the National Health Interview Survey. Due to their sensitive nature, much of the methodological research on administering this question has focused on survey mode, finding that self-administration reduces item nonresponse (Groves et al. 2004). However, self-administration may not always be an option, especially when instruments are complex or target respondents have lower levels of literacy. The research presented here explores another option for telephone administration of sensitive questions. This experiment uses the Worker Classification survey, a dual-frame nationally-representative RDD survey of 10,000 workers, to test two approaches to asking sexual orientation questions. We randomly assigned a sample of respondents to receive one of two versions of the sexual orientation question. In one version, interviewers read a response number followed by the response category (i.e., “one for gay, two for straight, or three for bisexual”). In the other version, interviewers read the category only (i.e., “gay, straight, or bisexual”). We hypothesized that offering numbered response options would result in lower item-nonresponse by adding a layer of privacy and potentially reducing the threat of disclosure. This is particularly relevant for respondents in the cell phone frame who may be more likely to complete the survey in the presence of others (Link et al. 2014). Our analysis compares item-response rates by experimental condition. We also explore possible differential response by demographic categories such as race, Hispanic origin, age, education, and income. Results will inform recommendations for future telephone administration of sexual orientation and other sensitive questions. Groves, R., F. Fowler Jr., M. Couper, J. Lepkowski, E. Singer, and R. Tourangeau. 2004. *Survey Methodology*. Hoboken, NJ: John Wiley. Link, M. W. et al. “Mobile Technologies for Conducting, Augmenting and Potentially Replacing Surveys,” http://www.aapor.org/Mobile_Technologies_Task_Force_Report.htm

Questionnaire Length and Response Rates: A Nationwide Experiment Across Three Modes of Administration

Rene Bautista, *NORC at the University of Chicago*

Reem Ghandour, *Maternal and Child Health Bureau at the Health Resources and Services Administration*

Alyson Croen, *NORC at the University of Chicago*

The effect of questionnaire length on survey participation is one of the most concerning topics among survey researchers; nonetheless, little empirical research has been conducted in the area. Typically, researchers work under the assumption that shorter questionnaires yield higher levels of response yet empirical evidence shows equivocal results. Since numerous factors are at play in survey participation—including mode of administration—it has been hard to estimate the relationship between questionnaire length and response rates, especially in the context of larger surveys. This paper will present results of a nationwide experiment on questionnaire length conducted in three modes of administration (i.e., CATI, web and paper) using a redesigned questionnaire of the National Survey of Children's Health (NSCH). The NSCH elicits data on children both with and without special health care needs. NSCH has been historically fielded as a telephone survey, but is looking to transition to self-administered modes (i.e., web and mail). As part of the redesign process, the NSCH instrument is combined with National

AAPOR 70th Annual Conference

Survey of Children with Special Health Care Needs (NS-CSHCN). Understandably, merging two questionnaires (NSCH and NS-CSHCN) produces a larger single questionnaire. To assess the impact of questionnaire length, NORC at the University of Chicago will conduct an experiment on behalf of the Maternal and Child Health Bureau (MCHB) at the Health Resources and Services Administration (HRSA) to test two versions of the redesigned questionnaire (i.e., long vs. short) across three modes of data collection (CATI, web and mail). Interview completion rates and response rates will be key metrics compared across conditions. Results and implications from this experiment using a representative sample of the U.S. population will be discussed in the paper.

Potential Applications of Text Analytics and Math-Based Coding Strategies for Write-In Responses

Tina Norris, *Kent State University*

Casey Tesfaye, *The Nielsen Company*

Jack Pold, *American Institute of Physics*

One persistent challenge in survey research is the coding of write-in responses to surveys. This can be a particular challenge when dealing with surveys that are repeated on a regular schedule. We strive to find coding solutions for these responses that are consistent over time and use staff time efficiently. In this paper we address the potential for text analytics and logic based programming statements as coding strategies. We focus on answers from two survey questions; 1. write-in responses to the Census race question and 2. the employer names provided in a follow-up survey of physics degree recipients. In both cases, either text analytics or math-based coding strategies were used to perform contextually accurate analyses as a means to clean data. Both of these coding efforts would have been significantly less efficient or consistent had they been accomplished through traditional string search or match programs. These cleaning strategies provided both programs a scalable mechanism by which to clean expansive volumes of data affordably by searching through and making sense of data in nearly real time. We will discuss the strategies we used, the methods we used to test these strategies and the applicability of the resulting code between sets of survey responses to the same question. As big data grows even bigger, new techniques in text analytics are growing in significance and practicality as a means to increase the accuracy and efficiency of overall data analytics. Incorporating text analytics is one mechanism by which federal agencies and private organizations can quickly see meaningful improvement in data quality while at the same time potentially reducing overall costs.

Can a Follow-Up Shortened Survey Increase Response Without Impacting Data Quality?

Jocelyn Newsome, *Westat*

Stephanie Beauvais Dennig, *Westat*

Kerry Levin, *Westat*

Brenda Schafer, *Internal Revenue Service*

Pat Langetieg, *Internal Revenue Service*

Ahmad Qadri, *Internal Revenue Service*

Ron Hodge, *Internal Revenue Service*

In the face of rapidly declining response rates, researchers have been exploring whether shortened surveys can be a method to reduce nonresponse bias without compromising data quality. Although some studies have shown lengthier surveys encourage satisficing, item non-

AAPOR 70th Annual Conference

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. In this study, we explore whether using a follow-up shortened survey with non-respondents increases survey response in an IRS survey without negatively impacting data quality. The IRS Individual Taxpayer Burden (ITB) Survey measures the time and money respondents spend to comply with tax filing regulations. It is conducted annually with about 20,000 respondents. The survey is comprised of 2 critical items that ask directly about time and money, along with 21 other items that provide context to respondents by asking more generally about the tax filing process. The shortened version includes only the time and money items and eliminates the contextual items. For the 2013 ITB Survey, we conducted an experiment where half of non-respondents received the original, 23-item “long” version of the survey and half received the “short” version for the sixth contact (3rd survey package mailing). In this paper, we examine the short version’s impact on overall response rates, as well as the impact on specific populations that have been historically underrepresented in the survey (e.g., younger adults, low income respondents, and parents with young children). We also assess the impact of the shortened version on data quality. In particular, we are interested in whether removing the contextual items results in respondents giving higher or lower estimates of their time and money burdens.

Changing Unbalanced Scales to Balanced Scales in the Peace Corps Annual Volunteer Survey: Experimental Study Results

Marina Murray, *Peace Corps*

From May 12 to May 28, 2014 an experiment was conducted to identify and quantify the effects of changing response scales from an unbalanced (also called asymmetrical or unipolar) to a balanced (also called symmetrical or bipolar) format in the Peace Corps Annual Volunteer Survey (a periodic survey of currently serving volunteers that was first administered in 1975). Two questionnaires with identical items but different types of scales in 58 of 206 items (20 of 55 survey questions) were administered online to a group of opt-in participants. A balanced or unbalanced version of the survey was randomly assigned to participants in week one. In week two, a respective second version followed to control for the order in which surveys are taken. Respondents were asked to allow approximately one week between the two surveys. Although 43 respondents completed the balanced version of the questionnaire and 40 completed the unbalanced version, the experimental results are based on responses from 33 participants whose balanced and unbalanced surveys were matched to control for possible effects of the respondents’ demographics. Statistical characteristics of scales, including inter-item interaction and discriminative ability, were compared between the two surveys. A change of scales from an unbalanced to a balanced format consistently resulted in more positive average ratings; however, a minimal impact on item discriminative ability and inter-item relationships was observed. This paper presents detailed results of the experiment with a focus on likelihood, effectiveness, and satisfaction scales. It concludes with a discussion about implications of scale changes for year-to-year comparisons in recurring cross-sectional surveys.

Methodological Challenges and Strategies for Veteran and Active Duty Military Survey Research

Methodological Challenges and Strategies for Veteran and Active Duty Military Survey Research

Heather Hammer, *Abt SRBI*

Erin Dursa, *Department of Veterans Affairs*

Aaron I. Schneiderman, *Department of Veterans Affairs*

Stanislav Kolenikov, *Abt Associates*

Surveys of veterans and military populations provide program planners, policymakers, public health practitioners, and academic researchers with an important source of epidemiological data on health-related outcomes, attitudes, and behaviors, exposures and confounders. Many are required to produce valid and reliable self-reported estimates of complex and often stigmatized mental health conditions such as post-traumatic stress disorder (PTSD) symptoms and other mental and physical health problems such as suicidal ideation, substance abuse, and irritable bowel syndrome, for example. Survey data are also used to identify associated risk and protective factors and health service needs. Estimates from these surveys are sensitive to the influence of question wording, mode, and context; frame coverage, nonresponse, and attrition; and the validation, weighting, and nonresponse bias study and adjustment procedures used. Although veteran and military populations are often described as highly-mobile, each branch, component, era, and operation presents both common and unique survey research challenges. This panel uses recently completed survey research on active duty military, reserve, and veteran populations to highlight some key methodological and ethical challenges and describe how they were addressed. The challenges include measurement; confidentiality; mode effects; location, recruitment and consent procedures; attrition, and nonresponse bias. Three of the surveys are longitudinal studies: the Follow-up Study of a National Cohort of Gulf War and Gulf War Era Veterans; the National Vietnam Veterans Longitudinal Study (NVVLS), and the Ohio National Guard Study.

Nonresponse Analysis and Adjustment in the Follow-Up Study of a National Cohort of Gulf War And Gulf War Era Veterans (Wave 3)

Heather Hammer, *Abt SRBI*

Erin Dursa, *Department of Veterans Affairs*

Aaron I. Schneiderman, *Department of Veterans Affairs*

Stanislav Kolenikov, *Abt Associates*

The Follow-up Study of a National Cohort of Gulf War and Gulf War Era Veterans is a multimode web, mail, and CATI survey. The original cohort for this longitudinal survey was comprised of 15,000 deployed Gulf War Veterans and 15,000 non-deployed Gulf War Era Veterans surveyed in 1995-1997 and then again in 2005. For the Wave 3 follow-up survey conducted in 2013, we used response propensity modeling (multiple logistic regression analysis) to examine the nonresponse (attrition) mechanism. Our assessment shows that the nonresponse weight adjustment via calibration successfully removed about 80% of the nonresponse bias in the test variable - marital status in 1991 as recorded in the frame data. It also significantly reduced the nonresponse bias initially observed in three of the five key Wave 3 survey variables examined: Unexplained Multisymptom Illness, smoking, and alcohol use. However, as evidenced by significant correlations between the examined outcomes and response propensities, we were unable to reduce the nonresponse bias for two of the five key

AAPOR 70th Annual Conference

variables: PTSD symptoms and general health. The negative correlation of PTSD symptoms with response propensity indicates that respondents tend to have lower PTSD screening scores (i.e., screen negative for PTSD) compared with nonrespondents. The positive correlation of response propensity with general health indicates that respondents tend to have higher self-reported general health compared with nonrespondents. This paper will describe the methodology used in the nonresponse bias analysis and discuss the performance of the nonresponse correction and its meaning for the results.

The National Vietnam Veterans Longitudinal Study (NVVLS)

Nida Corry, *Abt Associates*

William Schlenger, *Abt Associates*

The National Vietnam Veterans Longitudinal Study (NVVLS) is a congressionally mandated 25 year follow-up study of service members who were deployed to the war, 'theater Veterans,' and a comparison group who served during the war era, but were not deployed, 'era Veterans' who originally participated in the landmark National Vietnam Veterans Readjustment Study (NVVRS) in 1987. The purpose of the NVVLS was to assess the long-term course of PTSD, relationship of PTSD with physical illness, risk factors for PTSD, and service utilization among Vietnam veterans. NVVLS employed a non-equivalent comparison group two-stage quasi-experimental design; a mail and telephone survey of all participants that included screening assessments for the primary outcomes; and a probability sample who participated in a clinical assessment clinical diagnostic telephone interview administered by licensed psychologists. Following a 25 year interval with no interim contact, the cohort was located through an address search using the secure consolidated credit information from the LexisNexis Credit Information Bureau and searches of two national death registries to identify deceased cases. Of the 1,839 living participants, 1,450 (78.8%) participated in at least one of the survey components and 55.0% of selected living Veterans participated in the clinical interview. Of the total 1,450 participants who completed either the Phase 1 mail or Phase 2 telephone survey, 152 (10.5%) were recruited in person by field representatives, increasing the overall response rate from 70.6% to 78.8%. This paper describes the methodological approaches and key findings of the NVVLS, as well as implications for future research. Because the study employs assessment at two points in time and a registry that records deaths in the only nationally representative sample of Vietnam Veteran cohorts, it is a critical resource for scientific and policy analyses for Vietnam Veterans, with policy relevance for Iraq and Afghanistan Veterans.

When Research Impels Action: Balancing Evidence, Environment and Ethics in Studies of U.S. Veterans

Aaron I. Schneiderman, *Department of Veterans Affairs*

Victoria Davey, *Department of Veterans Affairs*

Erin Dursa, *Department of Veterans Affairs*

Robert Bossarte, *Department of Veterans Affairs*

More than a decade of military engagement by the U.S. Armed Forces has resulted in a dramatic increase in the number of military personnel and Veterans requiring care for physical and mental health concerns. Recognition of the psychological toll of combat and reintegration has driven demand for mental health services within the Department Veterans Affairs (VA). Concurrently, concern about suicide in the military and Veteran population has driven public dialogue by the government and stakeholder communities. Researchers from the Office of Public Health (OPH) have responded by querying research participants about mental health symptoms through the use of standardized screening instruments. VA is responsible for the

AAPOR 70th Annual Conference

evaluation of public health and provision of health care; an unusual position with regard to measurement of social and psychological phenomena where there may be a clinical requirement to assess and intervene for physical or mental health problems. DVA/OPH has used a range of methods to respond to reports of emotional distress or potential intent to harm oneself or others: non-intrusive notification regarding availability of services (notification of the toll-free VA Crisis Line); and active telephonic outreach to survey participants identified by survey responses that might indicate emotional distress. This paper discusses the approaches taken by the DVA/OPH in responding to participants in population-based surveys who endorse items suggestive of emotional distress during completion of a self-administered or CATI survey. Issues under examination include the predictive ability of clinical instruments when used in a general population survey; the ethical ramifications of altering the nature of confidential communication that is the foundation of the relationship between the health survey researcher and the research participant; and finally the potential conflict for a health care organization conducting surveys on sensitive topics including suicide and interpersonal violence when individual intervention is not part of the research intent.

The Impact of Retirement on Attrition in Military Service Cohort Research

Mark Morgan, *Abt SRBI*

Sandro Galea, *Boston University*

Joseph Calabrese, *University Hospitals Case Medical Center*

Marijo Tamburrino, *University of Toledo*

Israel Liberzon, *University of Michigan*

Gregory Cohen, *Boston University*

Daniel Loew, *Abt SRBI*

Laura Sampson, *Boston University*

Attrition is a challenge in any cohort study, and longitudinal studies of military personnel may have special challenges as the population is disproportionately young and highly mobile. In addition to their younger age, there is a fairly high rate of turnover as personnel leave the Guard to return to full time civilian life. The Ohio National Guard Study is a cohort study of roughly 3,500 Ohio National Guard members exploring the risk and resilience factors that influence the development of mental health disorders post-deployment including PTSD. This cohort project, begun in 2008, will consist of up to 9 annual surveys for each participant through 2018. Over the study period a substantial proportion of the cohort will leave the service and this poses a challenge to the measurement of mental health trajectories over time. Our analysis will compare the cohort attrition rates among retired personnel to those who remain in the Guard. We will also compare the retirees who remain in the sample to those who are lost to attrition to assess potential non-response bias on both demographic and outcome variables. Among those who remain active in the cohort we will compare retirees to those soldiers who have remained in the Guard.

Contribution of Methodological Differences to Variations in Reported Military Suicide and Suicidal Ideation Rates

Alisha Creel, *Abt SRBI*

Mark Mattiko, *U.S. Coast Guard*

Stephen Axelrad, *Booz Allen Hamilton*

Suicide is a focal concern for those working to improve the health and well-being of military and Veteran populations. Reported rates of suicidal behavior and ideation in military populations

AAPOR 70th Annual Conference

vary across data sources. This paper reviews data collection methodology and suicide rates from several military-wide and service-specific sources and studies: (1) the Department of Defense Suicide Event Report, (2) the 2011 Health Related Behaviors Survey of Active Duty Military Personnel, (3) the Millennium Cohort Study, (4) the Army Land Combat Survey, (5) the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS), and (6) the Air Force Community Assessment. We assess differences in item wording, overall survey context, respondent population, sampling design and weighting, and data collection mode and other procedures between the data collection sources used. Variations in suicidal behaviors and ideation are largely a result of the type of data source – with administrative data showing lower rates than self-reported surveys – and item wording. We discuss the implications of our findings for measuring suicide in military populations and make recommendations for comparing military data sources and implementing questions in future surveys. We also will discuss implications for making comparisons between military and civilian populations and the additional methodological concerns that researchers must consider when making such comparisons.

Challenges in Surveying Wounded Warriors

Suzanne B. Lederer, *ICF International*

Bradford Booth, *ICF International*

John Kunz, *ICF International*

Ronaldo Iachan, *ICF International*

Congress charged the 2010-2014 DoD Recovering Warrior Task Force (RWTF) with examining the effectiveness of DoD programs and policies for the care, management, and transition of Wounded Warriors. A cornerstone of RWTF's research plan was its mixed methods assessment of Wounded Warriors' experiences and views regarding the resources available to meet their needs, which RWTF accomplished through venue-based convenience sampling. Specifically, RWTF gathered survey and focus group data from 583 Wounded Warriors at 35 military locations. The survey results, in combination with other data, were used to support RWTF's recommendations but not to estimate the responses of the Wounded Warrior population. By leveraging the 2012 Defense Manpower Data Center (DMDC) Status of Forces Surveys (SOFS), RWTF attempted to gather similar data from Wounded Warriors through probability sampling of the total force. The June 2012 Active Component (AC) and Reserve Component (RC) SOFS obtained completed surveys from 12,325 AC and 26,826 RC Service members from initial samples of 65,000 and 112,824, respectively. The presenters, who provided research support to RWTF, will examine challenges inherent in surveying the Wounded Warrior population, based on their experience with these surveys and their knowledge of other major surveys of this or related populations. Such challenges involve defining "Wounded Warrior;" accessing sampling frames; reaching eligible non-users of services; eliciting participation from a rare and hard to reach population; and protecting vulnerable human subjects. Next we will describe RWTF's non-probability approach to surveying Wounded Warriors, highlighting its fit-for-purpose utility given RWTF's objectives and constraints, and the challenges inherent in surveying this population. We will conclude with ideas for introducing rigor to RWTF's approach. To adequately meet the needs of Wounded Warriors, the military needs more effective means of surveying them. AAPOR can provide the platform for optimizing methods for surveying, and thus serving, this worthy population.

Thursday, May 14
3:00 p.m. – 3:00 p.m.
Demonstration Session 1

Use of Precinct-Level Historical Election Results Data as a Post-Survey Adjustment Technique in Political Polls

Nigel Adrian Ronald Henry, *Solution by Simulation*

Post-stratification is widely accepted as a method to ensure representative samples, but it is also increasingly recognized that socio-demographic reweighting alone may do little to correct for differences in response propensities, or may even add more uncertainty to the estimates. Registration-based sampling (RBS) designs - like address-based sampling (ABS) - allows for the use of contextual data at various levels of geographic granularity corresponding to a respondent's or non-respondent's address. This demonstration shows a technique to include a potentially powerful covariate of non-response as a post-survey adjustment: political behavior as measured by precinct-level historical election results. This is akin to poststratification on party self-identification, but uses a theoretically more powerful variable and requires less assumptions.

Thursday, May 14
3:00 p.m. – 4:00 p.m.
Poster Session 1

Respondent Driven Sampling with Online Recruitment and Adaptive Follow-Ups

Ronaldo Iachan, *ICF*
Naomi Freedner, *ICF*
Christian Evans, *ICFI*
Karen Trocki, *ARG*

This paper describes a methodology that ICF developed in collaboration with the Alcohol Research Group for on-line respondent driven sampling (RDS) sample recruitment and telephone data collection. The methodology includes novel methods for selecting both seeds for the RDS sample from a national sample, and a comparison sample matched to the seeds subsampled from the same national sample. We provide some background for the national sample survey, the National Alcohol Survey, and the design of the study of sexual minority women that is the subject of the paper. We discuss alternative methods for matching including stratified random subsampling and propensity score matching. The paper also describes the adaptive approaches planned for controlling the RDS sample network structure. These approaches will use network data and para-data to control network parameters such as chain growth, or chain length, and intra-chain correlations. We discuss the importance of these parameters, and their control, for effectiveness of the RDS sample and for ensuring the quality of the data. In particular, these methods help prevent any possible scams that might arise from the online nature of the coupon distribution and the associated incentives.

Using Paradata to Manage and Monitor Collection and Assess Strategy

Christian Bertrand, *Statistics Canada*

Over the past few years, 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 among other, a series of dashboards and reports that can be used by managers to monitor results, identify issues and assess strategy. Active management can be defined as a set of plans and tools to manage data collection while in progress. The purpose is to provide timely, topical and relevant information on survey performance and progress throughout collection, so that problems with collection are identified early and decisions about how to correct problems. Response rate is not the only indicators that should be used to monitor and assess data collection performance and progress. Instead, response rate should be used in conjunction with other measures such as survey productivity, cost and representativeness indicators to make the best use of data collection resources while taking into account the trade-off between quality, timeliness and cost. In practice, this tool is also used by collection managers in Responsive Collection Design (RCD) context to monitor and analyse 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 this 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 and Responsive Collection Design. This paper describes the tools and strategies used to actively manage surveys at Statistics Canada.

Are Tailored Outreach Efforts Too Costly? An Assessment of a Responsive Design Approach to Control Costs and Nonresponse Bias

David P. Getman, *Decision Information Resources, Inc.*

Sylvia Epps, *Decision Information Resources, Inc.*

Leslyn Hall, *Redstone Research*

Jo Anna Hunter, *MDRC*

Responsive design approaches to data collection are increasingly common within the survey research community, as they afford planned opportunities to modify aspects of the study design to reduce total survey error and control costs through analysis of paradata (Heeringa and Groves, 2006). Researchers are able identify characteristics of unresponsive sample, apply less passive modes of data collection, and tailor outreach methods that are more appealing to them. Researchers are also able to shift resources away from the more responsive categories to focus on sample members more resistant to completing. This practice raises important questions related to how the increased focus on unresponsive sample categories affects overall response rates and cost to complete ratios. For example, does a tailored approach to specific sample categories lead to an increase in response and, thereby reduce nonresponse bias; if so, does that increase come at a response rate decline in other categories; or raise the overall cost per complete ratio so significantly that the benefits of this approach are not worthwhile? These questions form the bases of our case study and proposed poster. Using paradata from a single mix-mode data collection effort of urban, low-income, high-school students and recent high-school graduates (Web-Telephone-In-person follow-up), we demonstrate how a responsive design approach helped control costs, extend data collection, and reach response rate targets while staying within budget with a particularly hard-to-reach population. Design modifications

AAPOR 70th Annual Conference

included adjusting our outreach to accommodate the contact information we received, incentive increases, on-campus completion sessions, individualized mailings, and accessing alumni events. Initial results indicate that shifting efforts toward unresponsive sample categories did raise the response rate in those categories without negatively impacting other categories. Although the cost per complete ratio did increase for unresponsive sample categories, shifting resources created cost savings that enabled an extension of the field effort overall.

A Mail Survey Experiment Using Gallup's Annual Crime Survey

Jeffrey M. Jones, *Gallup*

Lydia Saad, *Gallup*

Frank Newport, *Gallup*

Stephanie Kafka, *Gallup*

This paper adds to the growing body of research comparing telephone and mail surveys. In October 2014, Gallup fielded parallel versions of its annual Gallup Poll Social Series (GPSS) Crime survey, using two sampling frames and related data collection modes: a random-digit dial (RDD) dual-frame telephone sampling frame for a telephone survey, and an address-based sampling (ABS) frame for a mail survey. The mail survey included a split-sample experiment testing two don't know/refusal options: half of the mail surveys included an explicit don't know/refused option for each question; the other half received an instruction at the beginning of the survey saying they may skip any question. The results of this experiment provide important data about how changes in the mode (interviewer administered by phone vs. self-administered via mail) affect the distribution of data and the ability to trend the results. Sampled households for the mail survey were also randomly assigned to three treatment conditions to assess the impact of incentives on participation and the subsequent demographic characteristics of the sample. No incentive, a \$1 up-front cash incentive and a \$2 up-front cash incentive are compared. The paper will fully explore the findings around these experiments, including a comparison of response rates for the two surveys, sample composition, distribution of don't know/refusal responses, and the results themselves. Because other researchers have found that respondents are more honest about sensitive topics in a self-administered interview, such as a mail survey, than in an interviewer-administered survey, such as a telephone survey, we are going to pay special attention to whether respondents self-reports of crime victimization in this survey, are also sensitive to over- or under-reporting by mode.

Measuring Water Governance: The Potential Impacts of Politics on Public Perceptions

Michelle L. Edwards, *Texas Christian University*

Government agencies and universities have typically been used to improve survey response rates and to increase the representativeness of survey samples (Heberlein and Baumgartner 1978). However, since the 1970s, data from the General Social Survey have demonstrated growing distrust in science among conservatives (Gauchat 2012). Other societal events also reveal potential lack of support for social scientific research among certain subgroups of the population, such as the 2012 amendment proposed by the Republican-led House of Representatives to eliminate the American Community Survey. In spring 2012, I surveyed an address-based sample of Washington and Nebraska residents on water governance issues using two different university sponsors (Washington State University and the University of Nebraska-Lincoln). At the end of this instrument, I included an open-ended question that allowed respondents to provide any additional thoughts. While these types of survey questions are infrequently analyzed, they provide additional insights into respondents' perceptions of the

AAPOR 70th Annual Conference

legitimacy of the survey process. Preliminary results based on these data suggest that a number of residents expressed concerns about the legitimacy of surveys sponsored by scientists, government agencies, and/or educational institutions. Potential implications for measuring water governance also discussed.

Observed Differences Between Hispanic and Non-Hispanic White Public Opinion Affecting Transportation Policy in the Lone Star State

Chris L. Simek, *Texas A&M Transportation Institute*

Hispanics make up 17 percent of the U.S. population, with this number expected to increase to 30 percent by 2060. Previously conducted research suggests that Hispanic travel behavior is substantially different from the travel behavior of other ethnic groups. Their distinct travel behaviors, coupled with their cultural and political views suggest that Hispanic opinions of transportation policy may also be unique, relative to other demographic groups. This research uses data from the Texas Transportation Poll to investigate differences between two specific groups of respondents (Hispanics and non-Hispanic whites) regarding their opinion of transportation policy in Texas. The findings suggest that Hispanic mobility challenges (reduced auto ownership and increased transit dependency) and political views are likely strong influences on their perception of transportation policy. The data also suggests that Hispanics are strong supporters of transportation funding increases, with a specific interest in funding public transportation. Furthermore, Hispanics are more likely to be more supportive of funding mechanisms that promote clean energy or fuel source independence, and government taking a more significant role in resolving transportation issues.

Measuring a Mayor's First Year

Micheline Blum, *Baruch College, CUNY*

Douglas Muzzio, *Baruch College, CUNY*

Eugene Averkiou, *Baruch College, CUNY*

Baruch College Survey Research and NY1 are collaborating on a public opinion poll of New York City examining views of Mayor Bill de Blasio and his first year in office. As part of this poll we will unveil a new index and new measures of performance for the city administration. We expect to continue using this index throughout the Mayor's term and track the public's rating of the administration. We will also explore various aspects of the administration and of perceptions of the Mayor, and analyze the relationship between these measures and the Mayor's overall rating. Baruch will conduct this poll of a random sample of approximately 800 NYC adults (landline and cell phone) using live interviewers in January 2015. The poll will be conducted in English and Spanish.

A New Look at the American Culture and Political Landscape

Daniel Cox, *Public Religion Research Institute*

Rachel Lienesch, *Public Religion Research Institute*

David Dutwin, *SSRS*

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

AAPOR 70th Annual Conference

address this shortage, Public Religion Research Institute (PRRI) has partnered with SSRS to launch the American Values Atlas (AVA). 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 and political 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 findings from the 2014 AVA covering a wide range of topics, including religious affiliation, political identity and views about important public policy questions such as abortion, same-sex marriage and immigration. We will pay particularly close attention to state-to-state variation in attitudes on these important questions and examine state-level changes in religious identity. The poster, like the AVA project overall, provides a unique window into the political and religious diversity in the U.S.

Corporate Political Behavior in a Post Citizens United Democracy: Examining Public Opinion of Super PACs, Campaign Finance, and Brands that Donate to Political Elections

Heather LaMarre, *Temple University*

In a post Citizens United democracy, the almost \$4 billion dollars spent on U.S. 2014 congressional elections broke several midterm election campaign spending records (Seigel, 2014, Nov. 4; opensecrets.org). Meanwhile, changes in the political landscape and media environment make it difficult for companies to remain silent on national political and social issues (LaMarre, 2012). Recent examples of highly visible corporate political behavior can be found in Amazon, General Mills, and Starbucks' political support for marriage equality ballot initiatives (Marzilli, 2012). As corporations exercise their right to donate unlimited amounts to political action committees (see *Citizens United v. Federal Election Commission*, 2010; Dunbar, 2012), public discourse regarding corporate political behavior has intensified (Freidman, 2010). LaMarre's (2012) study found a public opinion backlash against Target brands' donation to an anti-gay marriage gubernatorial candidate, which resulted in organized protests, flash-mobs, and the delayed opening of two San Francisco stores. This raises questions regarding the influence of corporate political behavior (CPB) (e.g., campaign and super PAC donations) on public opinion towards campaign finance and super PACs. Additionally, questions emerge as to whether such donations affect opinions of the brands or the candidates/issues they support. Using survey data (N = 504) with an embedded manipulation (e.g., news of a popular brand donating to an anti-gay marriage super PAC, donating to a pro-gay marriage Super PAC, and control), public opinion regarding CPB, campaign finance, super PACs, and the donating brands was analyzed. Results suggest a negative relationship between corporate political donations and brand attitudes, as well as general discontent with campaign finance laws and super PACs. Additionally, public opinion toward CPB (generally) and brands that donate (specifically) was significantly more negative among those who held strong opinions for or against gay-marriage. Additional results and implications are included.

Public Opinion and "Reforming the Energy Vision" (REV) in New York State

Carla Jackson, *Abt SRBI*

Tracey DeSimone, *NYSERDA*

New York State is initiating a massive effort to "Reform the Energy Vision" (REV) to address changes in the energy industry, including technological innovation and increasing

AAPOR 70th Annual Conference

competitiveness of renewable energy, against a backdrop of aging infrastructure, extreme weather events (such as Super Storm Sandy), and grid security needs. The objective of REV is to reduce greenhouse gas emissions in New York State by 80 percent by 2050. This is the most massive overhaul of the electricity sector since deregulation in the 1990s and if successful, would use demand, rather than generation, as the state's primary energy resource. New electric tariffs will seek customer-sited resources such as electric vehicles, demand response, rooftop solar, and energy storage. This aggressive statewide initiative will be rolled out against a backdrop of consumer opinion about energy. To develop a baseline as this effort is undertaken, a dual-frame, RDD survey of 800 New York State residents was conducted in July and August 2014. The survey addressed awareness of electric choice, interest in renewable energy, understanding of energy billing, perceived value of electric services, knowledge of household electricity usage, and consumer engagement in energy decision-making within the state. Survey findings revealed a high level of consumer interest in the delivery of energy to their homes and also in renewable energy. This and other major findings will be explored from the perspective of understanding how the 21st century consumer views energy and energy-related issues, and whether public opinion supports the changes which are being undertaken. The implications of these findings will also be explored for other jurisdictions seeking to expand the use of renewable energy and customer engagement in energy usage: What are the challenges? Is greenhouse gas reduction a concern for consumers? How engaged do consumers want to be in their energy future?

For-Profit College Graduates: Profiting From Their Education?

Andrew B. Dugan, *Gallup*

Stephanie Marken Kafka, *Gallup*

In 2014, Gallup, Purdue University and the Lumina Foundation created a unique partnership to measure the college experiences and long-term outcomes of college graduates nationally. This representative sample of over 30,000 college graduates is one of the most holistic, broadest investigations into the college experience as well as the long-term outcomes of college graduates ever conducted. The college student body is increasingly changing—according to the National Center for Education Statistics, less than a third of today's college students are of normal college age (18-24 years old) and attending a traditional 4-year not-for-profit college or university. Instead, many students – especially older adults looking to obtain their heretofore foregone higher education – are pursuing college degrees at private non-selective, for-profit universities. Virtually unheard of only two decades ago, for-profit universities have doubled their graduation rate over the past 14 years, and nearly one in ten college graduate obtaining an undergraduate degree between 2011 and 2014 attended a for-profit university. Even as enrollment at for-profit universities swells, there is little research examining what kind of imprint these nascent institutions leave on their graduates. Indeed, the social value of for-profit universities has become a popular topic in the nation's political debate, with little data to orient the discussion. This presentation offers a remedy by examining for-profit graduates on a number of important dimensions, from psychological ones such as alumni attachment to their institution, to experiential items and educational support in college and, ultimately, life outcomes as measured by five critical elements of well-being (financial, physical, career/purpose, social and community), work interest and engagement in one's job. For-profit graduates will be compared to their non-profit graduate academic counterparts to better understand the profile of a for-profit graduate.

AAPOR 70th Annual Conference

Comparing Perspectives on Choice in K-12 Education

Paul DiPerna, *Friedman Foundation for Educational Choice*

Public opinion research on K-12 education topics has gradually increased in recent years, but understanding for state and local views remains elusive. The vast majority of this public opinion research report out national results, and with the exception of several well-known survey programs (Education Next/PEPG/KN; PDK/Gallup; AP/NORC), public releases sporadically report demographic or other subgroup results. National polls are less useful for considerations in K-12 education policymaking, which is overwhelmingly done at the state and local levels. National surveys provide helpful-but-blunt indicators for policymakers, interest groups, researchers, and news organizations. However, the findings tend to miss the variation of public opinion from state to state, and across subgroups. K-12 education is a highly decentralized policy area, and even with the proliferation of polling in recent years, there is still a sparse amount of survey research in this policy area conducted at the state and local levels. Our poster will have a special focus on K-12 education and school choice issues, examining both national and state-level survey data. Since July 2010, we have collected polling data on registered voters in 23 states. We also have conducted annual national polls since 2012. The poster will compare national and state results, emphasizing levels, margins, and intensities across five topics (right direction/wrong track; school type preference; charter schools; school vouchers; education savings accounts, aka ESAs). Also, we plan to highlight annual trends as well as compare demographic responses (school parent, party ID, income, community type, age, race/ethnicity) on these core survey items. The purpose of the poster is to foster more interest in polling on K-12 education issues and corresponding public policies, and point out the cultural and policy perspectives generated by national, state, and demographic differences.

Student Perception Surveys and Educator Evaluation: Considerations for States and Districts Choosing or Developing a Student Feedback Survey of Teaching Practices

Clarissa Steele, *University of Wisconsin-Madison*

Since the No Child Left Behind (NCLB) Act expired in 2007, the U.S. Department of Education has granted states waivers from certain provisions of NCLB. In order to meet the waiver requirements, many K-12 state education agencies have designed new educator evaluation systems in order to identify effective and ineffective teachers and to provide data on the distribution of teachers of different levels of effectiveness across state schools and particular student subgroups. Most systems are based on standardized test scores and observations of teacher practice, and some states and districts are including student surveys in their teacher evaluation scores. These surveys ask students how their teacher manages the classroom, teaches content, and interacts with students. Few researchers have reviewed student perception surveys, and this paper will begin the discussion that many states and districts need to have in order to choose a survey aligned with their evaluation system, or to develop their own survey that meets their state's and district's needs. This paper provides a number of survey characteristics for states and districts to consider in choosing or developing a student perception survey, including the original purpose of each survey, survey validation for use in teacher evaluations, the grade levels and constructs covered by each survey, and survey testing for reliability and validity. Because teacher evaluation systems are a work in progress, this information may be helpful to state education agencies and local school districts to determine how a student perception survey can add information to a teacher evaluation that other metrics might not measure, and how a survey aligns with the state or district definition of an effective educator.

Analyzing Open-Ended Survey Questions Using Unsupervised Learning Methods

Fang Wang, *NORC*

Edward Mulrow, *NORC*

We explore using unsupervised learning to classify open-ended survey question responses. Unsupervised learning methods such as topic modeling or k-means clustering can provide techniques for organizing, understanding and summarizing text data without using any manually labeled records as training data. It uses annotations to organize text and discover latent themes in documents without target attributes. By grouping similar responses together, we construct a class of “topics” which are described by sets of keywords and reduce the exploration of open ended text information to common categorical analysis. We analyze the open-ended survey answers in the “10 Years After 9/11 Survey” and the “National Cultural Building Survey” using the text clustering and the topic model respectively as two examples. We also visualized the topic model results by an interactive HTML page using the R package LDAvis.

Big Data vs. Panel Data: The Importance of Representative Panels for Big Data Calibration

Shelli Kashriel, *The Nielsen Company*

Lindsey Rabhan, *The Nielsen Company*

Oana Dan, *The Nielsen Company*

As the revolution of “Big Data” continues, researchers are starting to rely heavily on large and diverse – yet often demographically ambiguous - data sources to measure consumer attitudes and behavior. In particular when measuring online audiences through machine-recorded information such as cookies, the accuracy of big data remains in question. Measurement that does not require direct participation from respondents makes it difficult to identify the true demographics of the person(s) actually consuming the content or interacting with the digital device at a specific time. In a recent study examining the effect of computer-sharing on online audience measurement, Nielsen compared demographic information of Internet users recorded from cookie data, to demographic information of Internet users collected from Nielsen’s computer-metered panel. The panel consists of approximately 10,000 demographically balanced households, where individual household members are required to log into the meter whenever they use the computer. By capturing consumer behavior in real time and with direct confirmation from panelists themselves, this panel serves as a source of truth for online media measurement, against which we compare “big data” about online behavior among the same households. Results indicate that online activity data captured directly from panelists differs significantly from online activity data recorded passively via cookies. For example, 24% of the time the demographics of the panelist actually consuming content on the computer did not match the demographics attributed by cookie data. The differences were caused by switches in computer users (which were recorded by panelists but not by cookies) and varied by gender, age, and household size. Thus, calibrating non-panel based data with representative panel data significantly improves the accuracy of audience measurement. When relying on “Big Data” for volume, velocity and variety, reliable panel data should be used to calibrate demographic information.

AAPOR 70th Annual Conference

Assessing the Use of Cell-Wins to Screen Nonworking Cellphone Numbers in Ohio

Kimberly Peterson, *RTI International*

Marcus Berzofsky, *RTI International*

Bo Lu, *Ohio State University*

Caroline Blanton, *RTI International*

Lance Couzens, *RTI International*

Jamie Ridenhour, *RTI International*

Tim Sahr, *Ohio State University*

Robert Ashmead, *Ohio State University*

Amy Ferketich, *Ohio State University*

Thomas Duffy, *RTI International*

In recent years, most surveys have introduced a dual-frame component to their surveys with the inclusion of cellphone numbers. However, due to a number of challenges including the restrictions in how cellphone numbers can be dialed it can be more expensive to complete cellphone interviews and often for this reason the proportion of the total sample that comes from the cellphone frame has been kept artificially small (e.g., 25% of the sample) to reduce data collection costs. Conversely, as more people shift to cellphones this percentage needs to increase from both a representation standpoint and to minimize variance. For example, in the state of Ohio, in 2012, 52.6% of persons were cellphone only or cellphone mostly users (Blumberg, et. al., 2013). These persons also tend to be younger, more likely to have children, and minority (Lu, et. al., 2014). Sample vendors have recently made improvements to the cellphone frame to help improve performance and in this paper we investigate the effectiveness of the Cell-Wins service provided by Marketing Systems Group (M-S-G) to identify if a cellphone number is actively being used by a person at the time the flag is applied. If accurate, this flag could greatly reduce data collection costs by identifying nonworking numbers prior to data collection. This paper presents an assessment of this flag in the State of Ohio conducted for the 2015 Ohio Medicaid Assessment Survey. Our study presents results of the flag's accuracy when it is applied within two weeks of contact as well as when the associated number is called a month after the flag was appended.

A Comparison of Internet and Telephone Election Polls in the Illinois Gubernatorial Election

Kirby Goidel, *Texas A&M University*

Ashley Kirzinger, *University of Illinois - Springfield*

We compare internet and telephone election polls for the Illinois gubernatorial election conducted by the Survey Research Office at the University of Illinois - Springfield during October 2014. The results of the surveys mirrored other polls conducted during the time period and the internet survey closely approximated the telephone survey. For further validation, we also compare survey results to the 2014 Senate election won by incumbent Democrat Dick Durbin, to the generic statewide House vote, and the results of a statewide tax referendum. Finally, we consider how both surveys compare to 2014 exit poll data. Overall, we conclude that the internet survey did a reasonably good job of capturing voter preferences at a significantly lower cost than the telephone survey.

AAPOR 70th Annual Conference

Looking Ahead to 2016: An Analysis of the Exit Poll 2016 Questions

Gregory Holyk, *Langer Research Associates*

Gary Langer, *Langer Research Associates*

The 2014 Exit Poll included several questions on the national questionnaire related to the 2016 presidential election, including whether or not various potential candidates would make a "good president" and a horse race pitting Hillary Clinton against a generic GOP opponent. The results are notably divergent and space on the exit poll is a precious commodity, which begs the question: How informative is it to ask a midterm electorate questions about a presidential election using a paper and pencil questionnaire? This paper compares the exit poll results to similar questions from national telephone polls to examine the influence of survey mode and sample population (general population, registered voters, likely voters and actual voters). The paper looks at demographic group differences by both mode and sample. In addition to comparisons of the the marginal and crosstab data, regression analyses are used to determine the strongest predictors of judgments of presidential potential as well as preference in the hypothetical horse race.

Rally in Russia: National Pride Surges in 2014

Julie A. Ray, *Gallup*

Neli Esipova, *Gallup*

Dato Tsabutashvili, *Gallup*

Propelled by a groundswell of national pride after the Sochi Olympic Games and Russia's annexation of Crimea, Russian President Vladimir Putin's popularity soared to some of the highest levels that Gallup's World Poll has found in eight years. The surge suggests Putin has solidified his previously shaky support base, but Russians' faith was also renewed in nearly all aspects of their lives – not just inside the political sphere. For example, Russians' confidence in their military, their national government and honesty of elections rallied to record levels, but they also had never been more satisfied with their freedom, they felt women and children were treated with more respect, and they perceived less corruption. At the same time, Russians' approval of the leadership of the U.S. and the EU hit all-time lows – underscoring the widening rift between Russia and the West – as their approval of China's leadership soared to all-time highs. This paper will use Gallup World Poll data to examine how far reaching the rally effect in Russian public opinion was in 2014 – and how deeply within society it was felt -- and contrast Russians' attitudes on measures that Gallup has tracked for the past eight years. The paper will also discuss other recent rally effects on public opinion in the U.S. to put Russia's rally within scope.

Measuring Public Opinion in Tunisia

Anita Pugliese, *The Gallup Organization*

Travis Owen, *The Gallup Organization*

Given the increased difficulties of conducting face to face surveys in the Middle East and North Africa, as well as the rapid increase in telephone penetration, Gallup has been testing the potential of alternative data collection methods in parts of the region. Previous research indicates that total landline and mobile coverage in Tunisia is nearly 90%, making it a viable option for data collection. We will present our findings from concurrent studies conducted with both face to face and by telephone (mobile and landline). We will describe our approach to obtaining nationally representative samples for these modes, and will report differences in

AAPOR 70th Annual Conference

geographic coverage, respondent demographics (age, gender, education, employment status, and household size); and core survey responses. Gallup has been tracking public opinion in Tunisia annually since 2006, on personal perceptions of wellbeing, safety, confidence in the national government, and other key measures. We will show trends of public opinion in this pivotal country from pre-Arab Spring through the latest measurement just prior to the 2014 parliamentary elections.

Crisis in Iraq: A View from the Ground

Nina R. Sabarre, *D3 Systems, Inc.*

Alicia Boyd, *D3 Systems, Inc.*

In June 2014, the Islamic State of Iraq and Syria (ISIS) seized control of the city of Mosul in northern Iraq. While the fall of Iraq's second largest city to ISIS militants came as a surprise to many, the long standing ethno-sectarian divisions within the country prefaced the collapse of the Iraqi military when challenged by ISIS insurgents. This study examines Iraqi public opinion during the two years leading up to ISIS's rise in Iraq and sheds light on how differences in attitudes between ethno-sectarian groups may have set the stage for the rapid ISIS seizure of Mosul. While media reporting can provide timely information about events in Iraq, it is necessarily limited to the sources available at the time. Conversely, a sustained effort to collect data using rigorous probability sampling offers unique insight into Iraqi public opinion that is authentic and broadly representative of the populace. D3 Systems, Inc., in cooperation with KA Iraq and KA Belgium, has conducted surveys in Iraq for more than a decade. This paper utilizes survey data from D3's Iraqi Futures project, consisting of nearly 20,000 nationally representative face-to-face interviews collected from 2012 to 2014. The authors of this paper analyze differences between the primary ethno-sectarian groups in Iraq (Kurds, Sunni Arabs and Shia Arabs) with regard to the perceived legitimacy of Iraqi government and security institutions. Looking at over 100 variables, the authors use logistic regression models to isolate variables (such as: ethno-sectarian identity, opinions of how the government handles security crises, attitudes towards political parties and leaders, and projected versus reported voting behavior, etc.) in order to understand their influence on perceptions of legitimacy and how these divisive views may have led to the current crisis and will influence the future course of the country.

Inter-Community Relations and Security in Kenya

Alicia Boyd, *D3 Systems, Inc.*

Timothy Van Blarcom, *D3 Systems, Inc.*

The International Criminal Court (ICC) has charged the current President of Kenya Uhuru Kenyatta with crimes against humanity for his alleged involvement in violence after the 2007 election. The election violence killed more than 1,000 Kenyans and President Kenyatta is the first serving head of state to be tried by the ICC. At the same time, security in the Kenyan capital, Nairobi, and areas in Coast Province, have been affected by terrorist violence perpetrated by the Islamist al-Shabaab group. In November 2014, D3 Systems conducted a nationally representative survey of Kenyans in order to understand Kenyans' opinions of: the ICC case against President Kenyatta; attitudes towards the Kenyan President himself; relations between members of different ethnic groups and tribes; and Kenya's role in military operations in east Africa to combat al-Shabaab. Using the data collected in November, this study examines the opinions of 1,000 Kenyans age 18+ on current domestic political and security issues. The paper investigates the attitudes of Kenyans across a wide cross-section of demographic variables – including tribal/ethnic group identity, age, education level, income, and occupational

AAPOR 70th Annual Conference

status – to analyze the influence of these variables on relations between ethnic groups, views of the Kenyan President and the ICC, and the impact of sending Kenyan troops to Somalia.

In Search of "Yugo-Nostalgia": An Examination of Public Attitudes in Post-Yugoslavian Nations

Elizabeth Keating, *Gallup*

Neli Esipova, *Gallup*

The collapse of the Socialist Federal Republic of Yugoslavia (SFRY) beginning in 1991 ushered in more than a decade of political and ethnic violence not witnessed on the European continent since the Second World War. Now, less than two decades after deadly clashes between Serbian, Bosnian, Croatian, and Kosovar nationalists elicited fears of genocide, a relative calm has returned to the region. Social and economic ties have deepened among former bitter rivals, while proximity to the European Union holds the promise of even greater prosperity. Nevertheless, as citizens of post-Yugoslav republics work to forge new national identities, questions remain as to what was gained and what was lost during the years of civil war. Using nationally representative data from the seven states comprising the former Yugoslavia: Bosnia and Herzegovina, Croatia, Kosovo, Macedonia, Montenegro, Serbia, and Slovenia, this study will examine the effect that the break-up and ensuing civil war had on citizens' economic welfare and political identities. Early findings suggest that most citizens of the former Yugoslavia believe that their countries are now worse off than before the break-up. Possible factors that influence this could include the human cost of securing independence, the loss of familiar social institutions, the threat of ongoing violence, and political and economic hardship. Findings also suggest that the degree of "Yugo-nostalgia" experienced today may also vary based on nationality, age, education, and income. Results of this study will also be compared with the results of a similar question that was asked in the former USSR about the break-up of the Soviet Union.

Differences in Political Opinions Between Arab and Jewish Israelis

Travis Owen, *The Gallup Organization*

Anita Pugliese, *The Gallup Organization*

Arab citizens of Israeli, or Arab-Israelis, make up roughly 20% of the Israeli population, but have long been seen as distinct from Jewish Israelis. They do not serve in the military, they live in separate areas, and they largely vote for different political parties. But little has been said about how Arab-Israelis themselves think about their country or their place in it. Using Gallup data collected in Israel between 2009 and 2014, we will explore the differences in life satisfaction and standard of living between Arab-Israelis and Jewish Israelis as well as their opinions on Israeli leadership and the Israeli-Palestinian conflict. Arab-Israelis are substantially less likely than other Israelis to be employed full time for an employer and have consistently expressed dissatisfaction with their standard of living. Only a third of Arab-Israelis have expressed confidence in the Israeli military and around 23% have confidence in the national government, compared to nearly half of other Israelis. We will explore these and other data to demonstrate how Arab-Israelis feel largely marginalized from their own state. In addition, we will compare Arab-Israeli opinions with those collected in Palestine.

Bystander Effects: How Does the Presence of Others Affect Response Distributions and Data Quality in Africa?

Charles Q. Lau, *RTI International*

Curtiss Cobb, *Facebook*

Michael Corey, *Facebook*

Andrew Fiore, *Facebook*

Diana Greene, *RTI International*

Min K. Lieskovsky, *Facebook*

Emilia Peytcheva, *RTI International*

A common assumption is that interviewers administer in-person surveys in private—away from other people who might hear the interview or influence respondents' answers. Practical experience in developing countries, however, suggests that family, neighbors, and acquaintances are frequently present during interviews. These “bystanders” may be more prevalent in developing countries due to the lack of space, novelty of surveys, and social norms. Despite the potential of bystanders to introduce error, there is little research on bystanders in developing countries. Our paper aims to fill this gap by analyzing data from in-person CAPI surveys conducted in Ghana, Nigeria, and Uganda (n = 3,000 in each country). The surveys use national probability samples of people age 15-64 years old, and ask a mix of factual and attitudinal questions about technology adoption. We pursue two objectives. First, we quantify the frequency of spouses, children, parents, and others being present, and the extent to which each person assists the respondent in answering survey questions. We investigate how these interactions are gendered: for example, parents may be more likely to observe daughters compared to sons. Second, we will use multivariate analysis to study how bystanders affect responses to questions that may be socially desirable. For example, children may report using the Internet and social media less often when their parents are present. In addition, we investigate how bystanders affect three indicators of data quality: item non-response, selecting the same response category on a common scale, and primacy/recency effects. This analysis tests the hypothesis that bystanders increase satisficing behavior because bystanders distract respondents from the survey task. Through this descriptive analysis, we aim to provide practical guidance to survey practitioners about how to minimize bystander effects.

The Nasty (and Engaged) Citizen: Associations Between Media Use, Political Incivility and Political Participation

Chance York, *Kent State University*

Danielle Sarver Coombs, *Kent State University*

Kirby Goidel, *Texas A&M University*

When individuals see displays of elite political incivility in mass media, they tend to hold more negative attitudes toward candidates and campaigns; at the same time, viewing uncivil political acts is related to increased interest in candidates and campaigns (Brooks & Geer, 2007; Cappella & Jamieson, 1997; Mutz & Reeves, 2005). How, then, does watching portrayals of incivility in mass media effect individual political participation when such displays inspire both negative views of politics and interest in politics? In this paper, we explore competing explanations for these effects by testing a communication mediation model of political incivility in which news media consumption—especially consumption of “hostile” cable TV and Internet news (Sobieraj & Berry, 2011)—channels the effects of demographics, ideology, and social-structural factors on uncivil attitudes toward candidates and participatory responses (see Lee, Shah, & McLeod, 2013). Using a unique, representative survey of registered voters (n = 1,009),

AAPOR 70th Annual Conference

we demonstrate support for a communication mediation model of incivility. Specifically, we find respondents who frequently use Internet and cable news are more likely to hold uncivil attitudes toward candidates—measured using one-word responses to Barack Obama in 2008 and 2012—than are mainstream news users. Second, we find that respondents who hold uncivil attitudes toward Barack Obama report higher rates of participation, even after controlling for ideology, political party identification, and other relevant factors. Frequent displays of elite incivility on Internet and cable TV news appear to cultivate uncivil political attitudes; nevertheless, holding uncivil attitudes relates to the normatively desirable effect of increased participation.

High-Intensity Media Coverage and the Impact on Public Opinion: A Study of Media, Murder and Public Sentiments

Ashley N. Cano, *East Tennessee State University*

Shonetra Walker, *East Tennessee State University*

Kelly Foster, *East Tennessee State University, Applied Social Research Lab*

In 2014, a small community in Virginia became focused on the disappearance of a young, white, college female. Throughout the course of this time between her disappearance and the discovery of her body, the police identified several similar cases that they link to a given suspect – thus turning a news story about the disappearance of a college student into one of a serial killer stalking women in the community. This study will look at how public opinion can be impacted by high intensity media attention. This study will focus on public concern over murder, kidnapping, and violence against women prior to and following this high profile case and the subsequent linking of a serial killer. Our study will utilize content analysis to determine public sentiment by examining volume of news stories, online comments and social media attention to see if the wording and language in their opinions changes at several points. The first point will be prior to the student missing, at the point the student goes missing, at the point a suspect is identified, and when the body is found. This will be accomplished by looking for common words and phrases used by the public and measuring the intensity and frequency of those words. Additionally, we will look at network-specific media bias towards person of interest versus the victim and the influence that has on public opinion of the case and whether or not media outlets frame the details of the case in ways that may skew public opinion. It is our hope that this project will showcase the ways in which media attention and social media can frame community dialog and impact people's perception about their community and about important social issues.

Archive as Dataset: Using the Roper Center Archive to Understand Polling and Public Opinion

Kathleen J. Weldon, *Roper Center for Public Opinion Research*

Elise M. Dunham, *Roper Center for Public Opinion Research*

Malaena Taylor, *Roper Center for Public Opinion Research*

The public opinion research collections archived at the Roper Center for Public Opinion Research have long served as a central resource for the polling community. While the availability of individual datasets and survey questions is of enduring value to researchers, the existence of a broadly comprehensive collection of major polling data in one place allows for innovative forms of analysis that cannot be conducted on a single survey or even the work of a single organization. Drawing on methodologies from media studies and text analytics, this poster will provide examples of new uses of the archive collection to reveal changes both in public perceptions and in the field of polling. We utilize the following approaches: tracking increases and decreases in polling on particular subjects; comparing the popularity of varying terminologies to describe similar concepts; marking the first appearances of particular topics

AAPOR 70th Annual Conference

and terms; and tracking changes in the number of surveys using particular methodologies and samples. The findings show that these techniques effectively reveal how pollsters themselves, through the questions they ask and the structure of the polls they conduct, both reflect and frame the parameters of public opinion.

A Systematic Generation of an Email Pool for Web Surveys

Henning Silber, *Göttingen University, Center of Methods in Social Sciences*

Jürgen Leibod, *Göttingen University, Center of Methods in Social Sciences*

Julia Lischewski, *Göttingen University, Center of Methods in Social Sciences*

Stephan Schlosser, *Göttingen University, Center of Methods in Social Sciences*

A constantly increasing amount of researchers are deciding to conduct online surveys. In terms of data quality, the data obtained by web surveys have been found to be of poorer quality than the data collected through other modes of data collection such as telephone and face to face surveys. The main critique of online surveys refers to the non-probability sampling methods in use. Whereas, it is comparably easy to draw random population samples based on the random generation of telephone numbers, online surveys are lacking this ability so far. Therefore, our research explores a new way of systematically generating a pool of e-mail addresses for online surveys. To this end, we applied a name based approach which is usually used in telephone survey research. In particular, we generated all possible combinations of the twenty-five most common German surnames, first names, and different e-mail providers. As a result, we produced a large pool of more than 400,000 e-mail addresses from which we could draw a random sample. To compare the results, we conducted the same survey among two samples from different online access panels and a student sample with the identical questionnaires and equivalent field times. All together, we are able compare four online surveys (N=100-300). This enables us to investigate a variety of aspects of data quality such as non-response, drop-outs and response time. Strikingly, the data obtained by our new sampling method performed well in all aspects of data quality and we would like to suggest it as a new possibility for online researchers to contact fresh respondents and ensure a wide spread among internet users.

Combining Statewide BRFSS Data to Produce National Prevalence Estimates

Kristie M. Healey, *ICF International*

Ronaldo Iachan, *ICF International*

The Behavioral Risk Factor Surveillance System (BRFSS) is the world's largest ongoing telephone health survey system. It is conducted as an independent study in each of the 50 states and the District of Columbia while adhering to guidelines set by the Centers for Disease Control (CDC). Although the primary purpose of the BRFSS is to produce reliable state-level estimates of major health risk behaviors, there is considerable interest in combining statewide data to produce national estimates. Similar methodologies and the universal use of a core questionnaire make data combination feasible. However, there are enough differences between states that combining unadjusted data could result in biased estimates. In particular, differential sampling rates across states yield high design effects when state data files are combined. This paper assesses the design effects and the level of bias in BRFSS estimates when combining unadjusted state-level data. It also introduces several new national-level weights with the objective of reducing bias and/or variability in estimates. Each new weight recalibrates the existing design weight to various combinations of national population totals. Estimates produced using the unadjusted weights as well as each newly developed weight are compared to national benchmarks from the National Health Interview Survey (NHIS). Preliminary results suggest that bias for six major health indicators is low when using unadjusted weights, and some reductions

AAPOR 70th Annual Conference

in weight variability can be achieved by recalibrating design weights. Additional efforts to reduce variability, such as weight trimming and subsampling states with high sampling rates, will be considered as well.

Out, Out, Damn Duplicates!

Marc I. Roemer, *Agency for Healthcare Research and Quality*

The Medical Expenditure Panel Survey (MEPS) has a follow-back component, the Medical Provider Component (MPC) which aims to increase the accuracy and completeness of the information collected in the Household Component (HC). Medical providers are contacted using the phone number, name, and address information collected in the HC. At times the HC inadvertently collects duplicative information about providers which in turn causes duplication in the MPC sample frame. This duplication of medical providers leads to the same point of contact receiving repeated calls from MPC data collection staff. Excessive requests for data raise burden, lower efficiency of data collection, and may cause deterioration in response rates. Besides these undesirable effects, a medical provider duplicated in the sample and contacted more than once may report the same data or different data each time; the resulting duplicated medical event data must be cleaned up after data collection has concluded, compounding inefficiency. For these reasons, we aim to eliminate duplication of medical providers in the sample frame of the MPC. Unduplicating providers does entail some risks and trade-offs. Excessive unduplicating risks failure to contact providers and acquire medical event data: a medical provider falsely grouped in a duplicate set will not report medical events which occurred at other medical providers in the set. A robust, effective method of unduplication must minimize false duplicates while maximizing the identification of true duplicates. This poster describes a method to unduplicate medical providers reported in the Household Component of MEPS before sampling and fielding the Medical Provider Component.

Are Some Interviewer Actions Increasing Interview Length?

Rebecca Gatward, *University of Michigan*

Piotr Dworak, *University of Michigan*

Discussion and analyses of the effects of interview length often focus on questionnaire length, design and interviewer behaviors or characteristics. Other specific actions an interviewer performs during an interview receive less attention. These actions can include; reference to question level help or instructions and the ability to enter interviewer notes to provide context or supplementary notes about data entered at a question. The amount of time interviewers spend performing these tasks varies due to the complexity of the questions, their experience level and may also just be due to habits the interviewer develops. For example, some interviewers may be providing an unnecessary amount of supplementary information in interviewer notes which do not affect data quality but do increase the length of the interview. We may also find that some interviewers refer more frequently on online help screens. This analysis will use paradata from the Health and Retirement Study to detail and explore the effect of this very specific set of interviewer actions on interview length. In addition we will discuss the optimum use of this functionality, which is now provided in computer assisted interviews to assist the interviewer and maintain data quality but could also be affecting interview length and thus respondent burden.

Changing ‘Who’ or ‘Where’: Implications for Data Quality in the American Time Use Survey

Caitlin E. Deal, *University of Nebraska, Lincoln; Survey Research and Methodology*

Antje Kirchner, *University of Nebraska, Lincoln; Survey Research and Methodology*

Ana Lucia Cordova Cazar, *University Of Nebraska, Lincoln; Survey Research and Methodology*

Lissandra Ellyne, *University of Nebraska, Lincoln; Survey Research and Methodology*

Robert Belli, *University of Nebraska, Lincoln; Survey Research and Methodology; Psychology*

Time diaries capture everyday life experiences in the form of a sequence of activities that occur in a given day and provide important insights into how time is used within our society. For valid inferences to be drawn, data must be as accurate as possible. The failure to accurately report certain aspects of activities is a large source of error for time diaries. Many studies have looked into associations between the accuracy of responses and certain respondent behaviors and found mixed results (Belli et al, 2013). Cognitive difficulty, which may be indicated by respondent uncertainty, such as replies of “don’t know,” longer response times, and inadequate answers, has been associated with less accurate data. However, in some cases these respondent behaviors have also been associated with increased accuracy. Indeed, when interviewers or respondents recognize these behaviors during an interview as cognitive difficulty and then work to overcome that problem, this can result in more accurate data. Paradata from the 2010 American Time Use Survey (ATUS) are used to investigate the relationship between response edits and the quality of the data. Specifically, this paper looks at whether changes in: (a) the respondents’ answers for their location during different activities and (b) who they were with during those activities are related to survey errors, such as memory gaps. Previous research has shown that the total number of ‘who’ and ‘where’ changes is not significantly related to the total number of insufficient detail errors or memory gap errors at the respondent level (Deal et al, 2014). Given that this lack of significance may be due to the aggregation of data from the activity level to the respondent level, this paper will explore whether a relationship may be uncovered using multilevel models with the activity level as the unit of analyses.

Who's Missing What on the American Community Survey?

Sandra Luckett Clark, *U.S. Census Bureau*

Nonresponse bias from questions left blank by respondents is present in all surveys. Measuring the amount of missing data, or item nonresponse, is an important indicator of data quality. The Census Bureau measures item nonresponse in the American Community Survey (ACS) to show data users the completeness of the data on which the survey estimates are based. We calculate and publish ACS item nonresponse rates for the entire population required to answer the items. However, these rates can vary for social, economic, and demographic subgroups, and by geographic area, which can lead to a greater nonresponse bias for some subgroups and areas than for others. This research uses 2013 ACS 1-year and 2009-2013 ACS 5-year data to examine item nonresponse rates for a variety of subgroups and small geographic areas. It includes aggregated rates for all population and housing items, along with individual rates for key survey items. The goal of the research is to determine how item nonresponse differs by geography and by demographic characteristics, such as age, race, citizenship, educational attainment, and income. The research will aid in our understanding of which areas and population groups have higher item nonresponse rates, which will provide data users and researchers needed information about the quality of survey data.

Covers and Grids – Two Questionnaire Design Experiments in a Mail Survey

Andrew R. Caporaso, *Westat*

Aaron Maitland, *Westat*

David Cantor, *Westat*

This report summarizes the results of two questionnaire design experiments embedded into the Health Information National Trends Survey (HINTS). HINTS is a cross-sectional mail survey sponsored by the National Cancer Institute which has historically achieved response rates between 35 and 40%. The survey takes roughly 30 minutes to complete and asks about health information seeking behavior, experiences with healthcare, cancer and other health outcomes. This HINTS cycle was fielded in the fall of 2014 and was sent to 14,000 households. The first experiment tested the impact of cover design on response rates and sample composition. Two design aspects were manipulated – background contrast (whether the cover background was white and the text was colored, or vice versa) and picture placement (whether pictures were grouped together or spread across the cover). We hypothesized that a non-white cover background would make the questionnaire more visible in respondents homes and therefore make it more likely that respondents see the questionnaire and fill it out. Picture placement was hypothesized to impact how much attention respondents pay to the cover and potentially impact their likelihood of responding. We also hypothesized that the cover pictures, which are of people engaged in health-related activities, could influence the types of people who respond to the survey. The second experiment tested the impact of grid shading on item-nonresponse in grid questions. Shading alternate items within a grid is a common tool for reducing missing data rates. This experiment will compare item-missing data rates in over a dozen grid questions between respondents whose questionnaires had shaded grids and those whose questionnaires did not. The final results and implications of both experiments will be presented and discussed. Preliminary results suggest modest differences in response rates across the covers and improvement in item-nonresponse for respondents who saw shaded grids.

Response Rates for Small Areas of Geography in the American Community Survey

Stephanie Baumgardner, *U.S. Census Bureau*

Response rates are a key measure of the quality of a survey, but it is important not just to look at the overall survey response rates, but at response rates for survey subgroups. The American Community Survey (ACS) has overall response rates at the national level of over 97 percent in a typical year, benefiting from the fact that it is a mandatory survey. However, the ACS publishes data for hundreds of thousands of geographic areas, so we need to also recognize the response rates for these smaller areas and understand how the rates vary by area. This paper examines the response rates for sub-state geographic areas such as counties, places, and Census tracts. We also analyze areas with relatively low response rates to identify the characteristics of low response areas. These characteristics include topics such as race, Hispanic origin, household size, owner/renter status, language spoken at home, education level, employment status, and poverty. Through this research, we learn about the types of areas that have lower response rates and provide researchers information about the severity of survey nonresponse across small geographic areas.

Measuring the Effects of Operational Designs on Response Rates and Nonresponse Bias

Kevin Ulrich, *University of Chicago Survey Lab*

Nora Henrikson, *Group Health Research Institute*

Melissa Anderson, *Group Health Research Institute*

Deborah King, *Group Health Research Institute*

As response rates continue to decline it is crucial to research areas that examine the conditions under which survey nonresponse results in nonresponse bias. Although there is a well-established body of literature that discuss methods to increase responses rates in web surveys, less is known about the effect of declining response rates and the link to nonresponse bias. The Snapshot survey, a representative survey of the membership of an integrated healthcare system in Washington state, included a randomized experiment testing various permutations of operational modes to measure the effects of nonresponse to a web-based survey. The use of administrative data from medical records allowed for the comparison of response rates, demographic characteristics, and health information of responders and non-responders. The experiment provided the opportunity to investigate multiple facets of nonresponse. Are the operational designs significantly different in terms of their resulting response rates, the demographics of the respondents and non-respondents and in the overall results? This paper will discuss the descriptive findings of the randomized experiments and discuss implications for future web-based health surveys.

Matching the Voice to the Voter: The Effects of Agent and Respondent Race in Automated (IVR) Surveys

Julie Martinez Ortega, *American Majority Policy Research*

Seth A. Rosenthal, *Yale Project on Climate Change Communication*

Matthew Fitch, *Merriman River Group*

Luke Henrici, *Merriman River Group*

We conducted a random-assignment split-half experiment comparing response and completion rates for a White call agent to response and completion rates for an African-American call agent. In a state-wide automated (IVR) election poll in Georgia ($n = 1,441$) late in the 2014 election cycle, half the respondents heard the survey instrument delivered by a White woman and half heard the instrument delivered by an African American woman. The women were approximately the same age, and read from identical instruments. We hypothesized that White respondents would be more likely to respond to the White voice than the African American voice. Results indicated that Whites responded to and completed the survey at significantly higher rates when presented by a White call agent relative to an African American call agent. Specifically, Whites were 49% more likely to begin the survey and 75% more likely to complete the survey in response to the White agent relative to the African American agent. Non-White respondents did not demonstrate this pattern. Notably, there were no significant differences in candidate preference patterns among Whites who responded to the White agent versus the African American agent. We surmise that this may have been the case, at least in part, because there were only White candidates, and no African American candidates, running in the races referenced in the survey. Overall, the magnitude of the differences in response and completion rates for the White agent relative to the African American agent indicates a strong preference for the White voice among White respondents in Georgia. This preference affected the demographic composition of the raw survey sample, but not the topline results of the key survey items after demographic adjustments were applied.

AAPOR 70th Annual Conference

Data Collection Challenges in Targeting Small Geographic Areas

Andrew M. Evans, *Abt SRBI, Inc.*

Amy S. Lightstone, *Los Angeles County Department of Public Health Health Assessment Unit*

Nicole C. Lee, *Abt SRBI, Inc.*

Tara Merry, *Abt SRBI, Inc.*

Capturing respondents residing in specific, often small geographic areas within counties or large cities can prove challenging. Obtaining the sample to target is one issue, while accurate and successful geocoding of responses is another. Regarding the latter, a further obstacle arises when respondents are unwilling to divulge their exact address. Thus, researchers are forced to find solutions to increase the willingness of respondents to provide this crucial information, and in lieu of that, determine other geographic location information that enable the placing of respondents in as precise an area as possible. We examine the issue using preliminary data from the 2014-15 Los Angeles County Health Survey (LACHS) and discuss different operational approaches used to increase respondents' willingness to provide an address. Specific areas of interest are Service Planning Area, Health District, and First 5 LA's Best Start Communities. Incentives were offered to cell-phone frame respondents to increase their willingness to participate, providing the benefit of obtaining their address. Additionally, for the landline frame, the wording of the address question was amended during data collection. Preliminary results indicate that changing the wording of the question to explain to respondents how they would benefit from sharing their address information increased their willingness to provide their address. Prior to the wording change, 29.2% of landline respondents provided their address. After the wording change, we observed a small, but significant increase to 35.6%. Among cell-phone respondents who were offered a \$10 incentive, 73.0% provided their address. We will discuss alternative approaches that can be used to place respondents in the geographic areas of interest when respondents are unwilling to share such personal information as their exact address. The results of the analysis present practical implications and solutions for future research involving the need of targeting small, discrete geographic areas.

Applying SAE Methods to Sporting Event Audience Measurement

Jiaquan Fan, *The Nielsen Company*

Etienne Josserand, *The Nielsen Company*

William Waldron, *The Nielsen Company*

Encompassing a large population with an ongoing panel can be very costly. Moreover, certain subsets of this population might deserve more attention than others. Yet, it is still important to measure the population with a sufficiently large sample to maintain nice properties of estimates (unbiasedness, reliability, etc.). One common solution is to have two different panels: one more costly with high quality, and another one which is cheaper and covers the rest of the population. This second sample can be improved by using small area estimation techniques as a consequence of the first high-quality sample. Nielsen measures radio listening across the United States, including audiences for sporting events. In the largest cities, electronic measurement is utilized to record exposure to games, with suitable sample sizes being maintained on an ongoing daily basis. In the numerous remaining smaller cities, traditional paper "diary" surveys are used to gauge a household's listening for a single week. A suitable number of diaries are completed to ensure reliable station audience estimates, at the seasonal level. However, there lacks ample sample sizes to accurately estimate events that occurred at an exact date/time in a specific week. We endeavor to answer the following: 1. Can the radio

AAPOR 70th Annual Conference

diaries actually be used to gauge event audiences that occurred at a specified date/time? Or do these surveys merely reflect a person's typical listening? 2. Using the metered data to construct a synthetic estimate for a particular game, can we build a viable small area model for single game estimates, as well as an estimate of the number of persons that listened to at least one game during the season (referred to as cumulative season audience)?

How Many is Too Many? An Examination of Call Design in Western Europe

Kenneth Kluch, *Gallup*

Sofia Kluch, *Gallup*

Bob Tortora, *Gallup*

Common call designs for CATI general population studies range from 3, 5 and N+ attempts. Multiple call attempts are used to best reach a targeted individual, provide a variety of calling times and dates to secure the correct respondent, and most critically, to reduce the potential for systematic bias introduced through non-response. Building on Pew 2012, and BLS 2009, this study examined respondents from 20 Western European countries reached on different call attempts across a variety of items and demographics to determine the representativeness of the sample and the ideal number of attempts required for a study. The Gallup World Poll data from 2014 Western Europe was used to examine impact of potential bias. In Western Europe Gallup uses a 5+5 call design where 5 call attempts are made on each number in the sample and if contact is made, an additional five attempts are made to secure a complete. In this analysis respondent were placed in one of three groups (1-3, 4-6 and 7 or more) based on the number of attempts required to complete the interview. While significant differences were noted for some countries between the groups, the analysis showed that in a majority of countries there were no differences in respondents based on the number of attempts. When differences were identified roughly half were between those reached on the first three attempts versus those reached on attempts 4-6. Few differences were identified between the later groups. In all countries studied 80% or more of completed interviews were recorded by the sixth attempt. With few difference between the latter groups and the high number of completes by the sixth attempt, the research suggests that longer call designs may not be needed in order to reduce potential bias in these specific countries.

Respondent Reporting Patterns for Questionnaire Items with Unknown Response Categories

Kimberly L. Ault, *RTI International*

Rachel Harter, *RTI International*

Jiantong Wang, *RTI International*

Patricia Green, *RTI International*

Peter Einaudi, *RTI International*

The Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS)—sponsored by the National Science Foundation (NSF) and the National Institutes of Health—is an annual survey of all academic institutions in the United States granting research-based master's degrees or doctorates in science, engineering, or selected health (SEH) fields. The GSS provides data on the number and characteristics of graduate students, postdoctoral appointees, and doctorate-holding non-faculty researchers in SEH fields. NSF uses the results of this survey to assess shifts in graduate enrollment and postdoc appointments and trends in financial support. In 2010, the GSS introduced new questions that include an unknown response category on primary funding source and doctorate type for postdoctoral appointees.

AAPOR 70th Annual Conference

This research determines the extent to which respondents use the unknown response category and investigates whether these responses could be allocated to other response categories through imputation. The analysis in this presentation is based on the 2010-2012 GSS data. The 2010 GSS collected data on the graduate students and postdocs from 13,711 organizational units (departments, programs, affiliated research centers, and health care facilities) at 574 institutions of higher education and their affiliates in the United States, Puerto Rico, and Guam. The institutional response rate was 99.3%. The 2011 GSS collected data from 565 core institutions with 686 schools and 13,793 units and the 2012 GSS included 565 core institutions with 684 schools and 13,952 units where the institutional response rate is over 99%. An investigation into the stability of response patterns over time assesses whether the previous year data carry forward imputation method is viable. An examination of correlations between imputation class variables shows support for either nearest neighbor or model-based imputation methods. Advantages and disadvantages for using imputation methods for allocation of unknown responses are discussed and proposed guidelines for implementation are presented.

Opportunity Survey: Understanding the Roots of Inequality

Eleni Delimpaltadaki Janis, *The Opportunity Agenda*

The Opportunity Agenda and Langer Research Associates conducted a survey to explore a range of key predictors of attitudes and behaviors toward social issues, such as poverty, criminal justice and racial discrimination. Some of the key ideological and psychological orientations explored in the survey include (1) the importance placed on key moral foundations, such as harm/care or fairness/reciprocity, (2) perceptions of outgroups, submission to authorities and aggressiveness towards outgroups, (3) Causal attributions for personal, ingroup and outgroup circumstances, as well as perceptions of deservingness, or (4) Personal and group efficacy (believing that taking action will make a difference) as motivational forces. Segmentation analysis and other appropriate statistical methods (regression analysis, factor analysis and/or structural equation modeling) will be used to identify the strongest independent predictors of a variety of outcomes, including but not limited to likelihood to participate in social movements and support for different policy proposals. We will also test models of interrelationships among constructs. The questionnaire design and data analysis were largely guided by a literature review, which examined existing work on the factors affecting attitudes and behaviors toward equality, opportunity and social issues, including ideological and psychological orientations associated with social justice. The final product is a 30-page report. The survey questionnaire is 20 minutes in length, adhering to best practices in questionnaire design. It is structured both to avoid bias and to facilitate statistical modeling. Administered by Langer Research Associates, the Opportunity Survey was conducted between February 4 and March 10, 2014, among a random national sample of 2,055 respondents. The survey oversampled very low-income adults (those living below 50 percent of the federal poverty line), African American men, and Asian Americans.

**Thursday, May 14
4:00 p.m. – 5:30 p.m.
Concurrent Session B**

Mini-Conference: Concerning Quality in Opt-In Panels

An Evaluation of Online Quality Control Questions

Keith Phillips Dustin Phillips, SS/

Many online survey writers incorporate questions to measure the attentiveness of the survey taker. The assumption is that if survey participants cannot accurately complete the quality control questions they are satisficing and should be removed from the study. However, the amount of participants who are providing “poor” data quality differs based on the design of the quality control measure. Yet, there are no industry standards or guidelines for the creation of these quality control measures. Most survey writers do not provide evidence that their quality evaluation improves the overall data quality of the study, but infer they work because they make logical sense. We interviewed 2100 online participants and tested 15 different quality control measurements for this study. We examined whether or not some quality control questions are more effective than others and the impact exclusion has on overall data quality. Are some of these questions evaluating quality of the participant or evaluating question design flaws? We look at online self-completion data against credible offline benchmarks and then compare our data set to these benchmarks as we remove participants based on their answers to different quality control questions. We also suggest proper evaluation techniques to remove “poorly” behaving participants. Our goal is to try to find the combination of satisficing checks that most accurately identifies bad behaviors. We do this by creating segments of individuals based on the number of quality control questions they failed. We then look at the results among these segments to see, where the data differs against the offline benchmark. Once we establish the bad behaviors we can then evaluate each quality control measurement in isolation to see which questions have the lowest false positive and false negative rates.

The Key Factor of Opinion Poll Quality

Shaw Tao, *Environics Analytics*

The inaccurate poll results of recent Canadian provincial elections incurred sharp criticism. Many studies have been conducted to find out the causes. As one of these efforts, this paper put its focus on one of the key factors affecting poll quality – data source. To remedy the listed telephone directory database decline, many data sources (or methods), such as cell phone RDD, online panels, social media, as well as traditional mail dwellings, have become more popularly used in market research and opinion polls. This paper provides a review and comparison of all these methods on key quality measurements such as coverage rate, response rate, bias error, respondent fatigue, barrier of responses, etc. Based on the conducted analyses, the author found that while being able to bring significant cost savings and fast turnaround time, some innovative methods are not mature enough to achieve reliable results.

AAPOR 70th Annual Conference

Applying New Technology to Global Population Insights Capture

Eric Meerkamper, *RIWI Corp*

The public opinion research field is undoubtedly experiencing a period of rapid change and innovation. The growth of mobile, nano-survey, social media analytics, and various other approaches is not only providing public opinion researchers with new ways of engaging existing respondent populations, but also new ways of engaging new global populations. RIWI will share findings from the 2014 GRIT Consumer Participation in Research report (to be released January 2015), a global study conducted in April-May 2014 using RIWI's Random Domain Intercept Technology (RDIT), which captured over 50,000 randomized, non-panel, non-incented respondents from 55 countries in 17 languages. Respondents were shown 20 questions, including opinions on political freedoms, online purchasing patterns, use of social media etc. About 70 percent of the respondents reported that they had not taken a survey of any type in the past month or longer, i.e. are "fresh responders", while the remaining 30 percent had answered a survey more recently, i.e. "frequent responders". On 19 of the 20 questions, there was a statistically significant difference between the average responses of the "fresh responders" and the average response of the "frequent responders". This presentation will provide attendees with new data and perspectives on:

- What are some considerations and challenges when engaging new global respondent populations?
- Are "Frequent Responders" different from "Fresh Responders"? If so, should this be of concern; and what are the implications for the industry and those who make decisions based on these data?
- What are the considerations when comparing or integrating insights from different populations? What is the potential impact on current longitudinal data sets?
- What are some of the strengths, limitations and most appropriate uses of emerging data capture technologies?

Exploring Causal Effects in Laboratory, Survey and Field Experiments with Nonrandom Study Participation

Jason Barabas, *Stony Brook University*

Jennifer Jerit, *Stony Brook University*

Carlos Paez, *Harvard University*

Many public opinion experiments use convenience samples with descriptively unrepresentative participants drawn from unknown populations. Some researchers argue that experiments with nonprobability samples—e.g., those using university undergraduates, online opt-in panels, or Amazon Mechanical Turk—are unproblematic and that causal effects generalize. However, often the evidence supporting these claims rests upon limited comparisons of demographic profiles, hypothetical simulations, or contrasting the results of "parallel" experiments—the results of which are ambiguous because subjects are selected and assigned to modes nonrandomly. We revisit the survey data collected in two published works and present new results from a third study that all utilize (A) simple random selection of subjects, and (B) random assignment to laboratory, survey, or field experimental interventions on a diverse set of political and public policy issues. Because all three of the empirical studies employ registration-based sampling (RBS) with diverse and well-defined populations with thousands of respondents drawn from statewide voter rolls, it means we can (1) document the existence of highly nonrandom participation in experiments, and (2) statistically adjust our results to account for selection biases. In doing so, we provide definitive evidence in support of the view that experimental effects are generalizable even if study participants do not mirror the target population.

Public Opinion Research in the 21st Century - Scientific Surveys Based on Incomplete Sampling Frames and High Rates of Nonresponse

Mansour Fahimi, *GfK*

Frances M. Barlas, *GfK*

Randall K. Thomas, *GfK*

Annie Weber, *GfK*

Traditional methods of survey research that rely on Neyman's probability-based sampling paradigm are based on a number of fundamental assumptions that may be unattainable under certain conditions. On the one hand, common methods of sampling are subject to coverage issues that may not be fully 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 need for cost containments and it is no wonder why alternative sampling methods are gaining popularity, even among traditional survey researchers. Top among such alternative methods are non-probability samples secured from online panels, commonly referred to as opt-in samples. Due to indeterminable selection probabilities and unknown coverage properties, however, measurable statistical inferences cannot be generated from such samples. The authors will review a number of practices that are currently used for developing inferences from opt-in samples, as well as discuss a robust weighting methodology that can further reduce the inherent biases associated with non-probability samples. Moreover, the relevance and applicability of this weighting methodology to probability-based sample surveys suffering from incomplete frames and high rates of nonresponse will also be discussed. Results are based on two midterm election surveys recently conducted in Georgia and Illinois; in each state two parallel surveys were administered using both a probability-based sample and a corresponding opt-in sample. The efficacy of our proposed methodology is assessed in light of comparisons of survey estimates to external benchmarks obtained from the American Community Survey and the Behavioral Risk Factor Surveillance System, as well as actual midterm election results in each state.

Polling in the 2014 Election: A Recap

Polling in the 2014 Election: A Recap

Dan Merkle, *ABC*

Anthony Salvanto, *CBS*

John Lapinski, *NBC*

Joe Lenski, *Edison Research*

The AAPOR Executive Council has dedicated this session to the memory of our friend and colleague, Mike Mokrzycki. Election polling has witnessed tests from two fronts. First, as the electorate has more quickly evolved, pollsters have had to ensure that their voter models are reflective of that cross section of Americans who actually show up at the polls. Second, survey research methods are evolving as well, with different emerging methodologies and even greater challenges in the exit polls. How did pollsters tackle these challenges in the 2014 midterm elections? What did they learn? This panel serves as a recap, straight from the experiences of the directors of polling at the three major television networks and the National Election Pool.

Measurement Error and Questionnaire Design

Assessing the Reliability of Measurement in the General Social Survey

Duane F. Alwin, *Penn State University*

Brett A. Beattie, *Penn State University*

Erin M. Baumgartner, *Penn State University*

This paper reports some preliminary results from an NSF-supported project concerning the estimation of the reliability of measurement in one of the most commonly used data sets in contemporary social science—the General Social Survey (GSS). Preliminary results are presented for nearly 500 questions from the GSS survey, applying the psychometric concept of reliability using the longitudinal approach to estimation, implemented by the GSS in two three-wave panel surveys—initiated in 2006 and 2008. Data were collected on representative samples of the U.S. population, with re-interviews implemented at two-year intervals. Estimates of reliability were obtained from both list-wise and full-samples, the latter using weighted least squares mean- and variance-adjusted (WLSMV) or full-information maximum-likelihood (FIML) to handle missing data due to attrition and other causes. Consistent with prior studies, list-wise and WLSMV/FIML estimates were virtually identical, suggesting an MCAR pattern to attrition and missing data. Detailed coding of question characteristics permits a meta-analytic approach to the analysis of attributes of questions on question-specific reliability estimates. Results are presented for measures derived from roughly 100 factual and 390 non-factual questions in the GSS panels, in which several hypotheses from prior studies are examined. This paper examines conclusions from prior studies concerning 1) survey content (fact vs. non-fact), (2) source of information (self-reports vs. proxy reports), 3) survey context (presence in series and batteries), and 4) attributes of questions (question form, number of response categories, labeling response options, explicit Don't Know options, and question length). Results generally support the conclusions of prior studies, with some important exceptions. The discussion of these results focuses on the general question of whether these and other results obtained in the estimation of reliability of measurement can be employed to improve the development of survey question strategies. Several areas are identified where specific recommendations can be made.

Item Nonresponse, Heaping and Response Certainty in Subjective Probability Questions

Sunghee Lee, *University of Michigan*

Florian Keusch, *University of Mannheim*

Colleen McClain, *University of Michigan*

Questions using subjective probability have become a popular format in survey research. Examples range from questions asking respondents their chances to vote in an upcoming election and to those asking the chances that the benefits of age-based entitlement program (e.g., Social Security in the U.S.) will change in the future. While the popularity is impressive, these questions require respondents not only to think about future events that may or may not be relevant to them and to provide the best guesses in a numeric format that ranges 0 to 100. In other words, respondents may need to engage in a higher level of cognitive efforts to answer these questions than other survey questions, which translates into potential item nonresponse and responses heaping at numbers that are multiples of 10 or 25. We will examine the measurement of subjective probability questions in the 2010 Health and Retirement Study (HRS) in the following three respects. First, item nonresponse and heaping patterns will be examined across a set of subjective probability questions, including subjective life expectancy,

AAPOR 70th Annual Conference

future out-of-pocket medical expenses, changes to the Social Security benefits, and an increase in the worth of mutual funds. Second, for respondents who reported a 50% probability, HRS 2010 asked a follow-up question probing whether they were uncertain about the chances or they believed that there were equal chances. Combining the data from the actual questions and the probing questions, we will examine the level of certainty in responses to subjective probability questions. Third, we will add another dimension to the analysis: actual interviewer-respondent interactions on these questions by behavioral coding audio-recorded interviews. In all analyses, we will examine the role of respondent characteristics (e.g., cognitive capability) on the response patterns and certainty.

Comparing Direct and Filtered Frequency Questions: Which Produces More Accurate Measurements?

Rajesh Srinivasan, *Gallup*

Annabel Suh, *Stanford University*

Jon Krosnick, *Stanford University*

A number of studies have shown that asking a direct question (e.g., How many movies did you watch?) yields reports of more events than does asking a pair of filtered questions (e.g., Did you watch a movie? IF YES: How many movies did you watch?). The only proposed explanation for this difference is the implication that a person asking a direct question presumes that the respondent has performed the behavior of interest and therefore causes inaccurately inflated reports to be provided. However, this explanation has been discredited by a number of direct tests. Thus, it is not clear which of the two question forms yields the most accurate results. This presentation will report the results of a new experiment designed to determine which question form yields more accurate reports. Data were collected in two waves from the same individuals. During the first wave, all respondents were asked a large number of questions on a variety of topics. During the second wave, respondents were asked to report how many questions they had been asked during Wave 1 on each of those different topics. Respondents were randomly assigned to 16 different treatment groups based on direct or filtered form, broad or specific topic, and time interval between wave 1 and wave 2 data collection. Because the correct answers to the recall questions are known, it is possible to determine which Wave 2 question format yielded the most accurate reports. Further, it is possible to explore whether the difference between the two question formats is a function of the difficulty of the recall task, as determined by the length of the recall period and the specificity of the topic. The results will provide a basis for recommending which question form should be used to measure event frequencies.

Priming Mindful Responding: Relationships Between Survey Instructions, Mindfulness and Data Quality in a Telephone Survey

Colleen A. McClain, *Michigan Program in Survey Methodology*

Florian Keusch, *University of Mannheim, Germany*

Ting Yan, *Westat*

David L. Vannette, *Stanford University*

James M. Lepkowski, *Michigan Program in Survey Methodology*

A large body of research investigates conditions under which satisficing (Krosnick, 1991) may lead respondents to take cognitive shortcuts, circumventing the optimal response process. The relevance of task characteristics and respondent motivation in the framework raises the possibility that researchers may be able to use survey instructions to frame items in ways that minimize satisficing. However, current literature does not fully address the effectiveness of such

AAPOR 70th Annual Conference

“primes,” or the conditions under which they may be most effective—especially when these conditions may be transient states rather than stable demographics. One condition that has been proposed as a linkage to the satisficing framework (Vannette & Krosnick, 2014), but not yet studied empirically in this context, is mindfulness (Langer & Abelson, 1972; Langer, 1992; Sternberg, 2000; Lau et al., 2006)—a multi-dimensional construct that involves conscious awareness, orientation in the present, and non-reliance on scripts, and that can be induced and changed via interventions. The present study expands upon this linkage using data from the June 2014 Supplement to the Surveys of Consumers (SCA). It assesses whether a survey instruction designed to prompt mindful processing and administered to a random half of respondents will impact respondent motivation and reduce satisficing. Further, it hypothesizes that these effects will be moderated by self-reported level of mindfulness as assessed prior to the experimental manipulation. Key proxies for satisficing in the rest of the instrument—completion time, acquiescence, and midpoint response style—are assessed by the experimental condition and mindfulness level using bivariate and multivariate analysis techniques. Preliminary results indicate some significant effects of the instruction and mindfulness level on completion time and acquiescence. They suggest important future directions for the exploration of the link between satisficing and mindfulness in order to better understand contextual, motivational aspects of optimal survey response.

The Effects of Social Distance on the Construct Validity of Proxy Responses

Paul J. Scanlon, *National Center for Health Statistics*

How do proxy respondents with varying levels of social distance—the familiarity one person has with another—between themselves and their subjects answer questions? Given not only the shifting survey landscape where proxy data might be necessary because of high non-response, but also the fact that certain surveys must, by nature be administered to proxies, an increased focus on the construct validity of proxy reporting is necessary. In particular, the effects of the relationship between the proxy subject and the respondent requires more study. While many surveys that use proxy reporting are designed to sample close family members, this is not always possible in the field where households might consist of a number of un-related individuals or where close family members are unavailable at the time of an interview. Thus, the individual who actually responds to a survey might have a dramatically different degree of social distance than the person for whom the survey was designed. Using cognitive interviewing and field data from a pair of collections—the 2014 WHO Verbal Autopsy Instrument and the UNICEF/Washington Group on Disability Statistics Module on Child Functioning and Disability—we will show that proxy respondents with varying levels of social distance both answer and interpret questions differently. Survey designers have the option of either attempting to strictly enforce their sampling preferences, or developing questions that can be answered validly by a range of proxy respondents. The second of these options is more realistic. Using cognitive data collected from these two projects, as well as from other evaluations of instruments that rely on proxy response, a number of design principals will be discussed. These include using framing to separate behavior and identity, as well as emphasizing a “don’t know” option in certain cases to reduce response error.

Measurement of Sexual/Gender Identity, Sexual Orientation and Same-Sex Relationship and Marital Status

Measurement of Sexual/Gender Identity, Sexual Orientation and Same-Sex Relationship and Marital Status

Kristen Miller, *National Center for Health Statistics*

James Dahlhamer, *National Center for Health Statistics*

A recent conversation on AAPORnet made clear the need for a practical conversation on the measurement of sexual/gender identity and sexual orientation. Our goals are to validly measure gender identity and sexual orientation in a manner that allows all respondents to feel their identity is adequately represented by the question and answer categories. To these ends, measures must provide respondents with an exhaustive list of response options that avoids forcing respondents into categories they may deem ill-fitting. While important, these objectives have proven to be difficult to put into practice. For example, extending the list of sexual/gender identities and sexual orientations can contribute to the cognitive burden of many in general population samples. To move toward these goals, this panel will discuss current and ongoing research on measurement of sexual/gender identity, sexual orientation, and relationship status for same-sex couples. Panelists will discuss findings from surveys ranging from federally sponsored, to state level, to local community-based organizations using data and paradata from pilots and pretests, full production surveys, and survey experiments. Papers 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.

Development and Resulting Data of a Sexual Identity Measure for the National Health Interview Survey

Kristen Miller, *National Center for Health Statistics*

James Dahlhamer, *National Center for Health Statistics*

For several decades, the lesbian, gay, bisexual, and transgender health movement has pushed for the collection of nationally representative health data of the LGBT population. Such data are required to achieve health equity, eliminate disparities, and improve the health of LGBT people. To this end, multiple health surveys have written their own 'sexual orientation' questions. The resulting data, however, has generated non-comparable estimates across surveys as well as high rates of missing data, particularly among low SES and Spanish speaking respondents. In reality, sexuality is a complex phenomenon that incorporates numerous, even contradictory, meanings, attitudes, and types of experiences which create a major challenge in developing a single measure that is both meaningful and accurate. The production of flawed data can lead to a mischaracterization of sexual minorities and, in the end, promote inadequate policy and health programs. It is critical, therefore, that a single quality measure be developed that most accurately collects data pertaining to this population. In 2013, the National Center for Health Statistics worked to develop and test a sexual identity question for the National Health Interview Survey (NHIS) that would account for the problematic complexities surrounding a single sexual identity question. Both qualitative and quantitative methods were used to develop and test a final measure. This paper will describe the development process for the measure. Additionally, 2013 NHIS data will be presented, illustrating how estimates are improved by the new design.

AAPOR 70th Annual Conference

Testing Sexual Orientation Questions in the National Survey on Drug Use and Health

Grace O'Neill, *SAMHSA*

Rachel Lipari, *SAMHSA*

David Dean, *SAMSHA*

The U.S. Department of Health and Human Services (HHS) intends to add sexual orientation questions to all HHS-funded surveys, where deemed appropriate, as they are redesigned. As part of the testing for the 2015 redesign of the National Survey on Drug Use and Health (NSDUH), two questions were added to the 2013 NSDUH Dress Rehearsal, one question collected information on sexual attraction and another on sexual identity. The inclusion of these items in the field test provides a limited opportunity to assess their impact on trends prior to adding them to the 2015 NSDUH. In this paper we discuss research conducted to identify which questions best fit the data needs of NSDUH, including initial placement and administration, the results of the NSDUH Dress Rehearsal, and the plans for placement of sexual orientation questions in future iterations of the survey.

Using Verbal Paradata Monitoring and Behavior Coding to Pilot Test Gender Identity Questions in the California Health Interview Survey: The Role of Qualitative and Quantitative Feedback

Matt Jans, *UCLA Center for Health Policy Research*

David Grant, *UCLA Center for Health Policy Research*

Royce Park, *UCLA Center for Health Policy Research*

Jane Kil, *UCLA Center for Health Policy Research*

Joe Viana, *UCLA School of Public Health*

Elaine Zahnd, *Public Health Institute*

Sue Holtby, *Public Health Institute*

This study reports on interview monitoring conducted as part of gender identity question testing in the California Health Interview Survey. Four methods of asking gender identity or transgender status were randomly assigned across approximately 3,000 respondents in production telephone interviewing. The first two methods used a two-question process, which first asked the respondent's sex assigned at birth followed by the sex/gender they identify with now, including an explicit "transgender" option. One of the two-question methods included explicit options "not sure yet" and "I don't know what this question means" options. The second two methods used one question, which began with a short transgender definition and asked the respondent if they identify as transgender. The one-question versions differed from each other in length and detail of the transgender definition. One of the one-question versions included a normalizing phrase, "some people describe themselves as transgender..." and a detailed definition that the interviewer could read as needed. Approximately 180 recordings were monitored (about 75% in English and 25% in Spanish, 50% with men, and a range of ages among those 18-70 years old). The monitoring form was designed to be easy to use in a quality assurance monitoring context, but also incorporate contemporary verbal paradata findings. It combined elements of traditional behavior coding (e.g., major and minor changes by the interviewer and requests for clarification by the respondent), paralinguistic verbal paradata, and other verbal behavior found to correlate with difficulty understanding and answering survey questions (e.g., fillers like "ums" and "uhs", answering too fast or too slow, non-answer verbalizations [i.e., reports], and mid-utterance answer repairs). We discuss theoretical and

AAPOR 70th Annual Conference

survey practice motivations for the monitoring form and process, and explain how monitoring results were used in conjunction with other pilot test information to evaluate the four methods.

One Question or Two? Measuring Gender Identity and Sexual Orientation Using a Mark-All-That-Applies Question Format

Justine Bulgar-Medina, *University of Massachusetts Boston*

Philip S. Brenner, *University of Massachusetts Boston*

Standard survey measures of gender identity and sexual orientation fail to capture approximately ten percent of respondents (Korchmaros et al. 2013). To improve measurement, best practices suggest more exhaustive but not necessarily exclusive response options (Williams Institute 2009). This approach, however, implies that a mark-all-that-apply list is a preferred way to capture these respondents. While lowering respondent burden relative to a series of forced choice items, mark-all-that-apply can be problematic. Current findings suggest that even for questions with high topic salience, satisficing may occur on mark-all-that-apply items, leading to reductions in data quality and requiring potentially unwarranted assumptions about item missing data (Smyth et al. 1996). Moreover, high topic salience can lead to verbatim responses to prior questions that would better fit the pre-coded categories of subsequent items. In a needs assessment survey for a Lesbian, Gay, Bisexual, Transgender, Queer and Ally (LGBTQA) community organization, a question wording experiment conducted online tested two versions of a question asking about sexual/gender identity and sexual orientation. Respondents received either (1) a single mark-all-that-apply question asking for both respondent's gender identity and sexual orientation, or (2) two separate questions on subsequent pages asking for the same information; one for gender identity and another for sexual orientation. Evidence for satisficing emerged in the one-question condition as many respondents chose a gender identity (a primacy effect) but not a sexual orientation. The two-question condition yielded verbatim responses with sexual orientation recorded offered in the "other" category to the first (gender identity) question. Findings suggest that, even for motivated respondents for whom the topic is of high salience, mark-all-that-apply can still promote measurement errors that are potentially harmful for data quality.

Putting the "T" in LGBT: A Transgender Question Pilot Test in the California Health Interview Survey

David Grant, *UCLA Center for Health Policy Research*

Matt Jans, *UCLA Center for Health Policy Research*

Royce Park, *UCLA Center for Health Policy Research*

Ninez Ponce, *UCLA Fielding School of Public Health*

As a large population-based health survey of the nation's most diverse state, the California Health Interview Survey (CHIS) has broken ground by making it possible to study many groups that are usually invisible or aggregated in other health surveys, such as Asian and Latino ethnic groups, linguistically-defined groups, and sexual minorities. Despite being the "T" in LGBT, transgender persons have been left out of this expansion. Nationwide, there are very few population-based estimates of transgender identity prevalence and they are often based on relatively small samples. The combination of low prevalence (estimated at less than 1%) and a complex concept (lived gender vs. biological sex) make these difficult questions to ask in general public health surveys. Indeed, gender identity may be a more complex question for respondents who have never considered the difference between assigned sex and self-identified gender than for transgender or gender-questioning respondents. In order to better understand the size of this population and its unique health characteristics, CHIS pilot tested

AAPOR 70th Annual Conference

four gender identity questions in the final quarter of CHIS 2013-2014. Questions were randomly assigned across approximately 3,000 respondents age 18 to 70. The pilot test was administered in English and Spanish and was conducted in production data collection, producing results that are direct estimates of the transgender population and a test of question wording effects. Two of the four questions asked about gender identity (i.e., a person's current identity as a man or woman) and their assigned sex at birth. The other two questions defined transgender and then directly asked respondents if they identified as transgender. We present differences in transgender identification, missing data rates, and interview breakoffs across the four questions. Our pilot test results expand best practices for telephone administration of questions to identify transgender adults in population health surveys.

It's So Simple, So Why Is It So Hard? Results from Tests to Reduce Measurement Error in Counting Same-Sex Couples

Nancy Bates, *U.S. Census Bureau*

Daphne Lofquist, *U.S. Census Bureau*

Jamie M. Lewis, *U.S. Census Bureau*

Matthew Streeter, *U.S. Census Bureau*

A recent Slate blog titled "Why Can't the Census Count Gay Couples Accurately?" laments the agency's admission that the estimate of gay married couples in the U.S. is almost certainly incorrect. Our paper examines this premise and reports on recent efforts to more accurately count this small subpopulation. Since the 2010 Census, the Census Bureau has attempted to quantify the degree of measurement error around traditional methods of enumerating same-sex couples. Following a program of qualitative research, the agency engaged in quantitative tests and experiments aimed to reduce the errors. The tests employed two methods designed to improve reporting. First, categories of relationship to the householder were expanded and revised to include "opposite-sex husband/wife/spouse," "opposite sex unmarried partner," "same-sex husband/wife/spouse", "same-sex unmarried partner ". Second, in the case of automated data collections, soft edits were added to flag and potentially correct reports where relationship appears incongruous with a couple's reported sex composition. Both are designed to reduce a tiny fraction of misreporting among the huge pool of heterosexual married couples in the US. Our paper reports findings from two tests – the interviewer-administered 2013 American Housing Survey (AHS) and the 2014 Census Test, a multi-mode test enumeration conducted in parts of the District of Columbia and Maryland. Both tests included split-panel experiments of the old versus new relationship questions. The AHS also included a soft pop-up edit alerting interviewers to inconsistent relationship/sex reports. In the paper, we report relationship item nonresponse rates, potential misclassification rates of reported same-sex couples, and the reduction of error resulting from the automated edit. We base these findings upon survey response data, survey paradata, and a name/sex probability index.

Improving Surveys with Usability Testing

Improving Surveys with Usability Testing

Emily McFarlane Geisen, *RTI International*

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. When usability testing is conducted, it is

AAPOR 70th Annual Conference

often conducted as an afterthought and not built into the survey development process. In this panel discussion, we will discuss recent usability testing projects that have revealed important findings that led to changing the appearance of the surveys and ultimately improved the user experience. Attendees will learn about the importance of conducting usability testing during survey development.

Making Usability-Testing a Standard Survey Pretesting Methodology

Emily McFarlane Geisen, *RTI International*

The capabilities of web surveys – particularly mobile web surveys – are constantly emerging. Examples are the use of cameras, videos, GPS and interactive features. As a result the need to evaluate how users interact with these products and how the experience affects the quality of survey data collected continues to grow. Despite this growing need for usability testing, it has not yet become a common or standard pretesting methodology for many web surveys. When usability testing is conducted, it is often conducted as an afterthought and not built into the survey development process. Usability testing results are often not published because they are conducted with small groups of participants and results are often specific to the survey tested and are not necessarily generalizable. When usability findings are published, the focus is usually on the results and with little or no discussion of the methods. Consequently, there is very little documentation on best practices for conducting usability testing on surveys. The purpose of this presentation is to present an iterative and design-based method for conducting usability testing that researchers can incorporate into the survey development process. I will provide examples of how usability testing can be applied at different stages in the survey development process from early prototypes and wireframes to fully programmed surveys. In addition, I will discuss how usability testing practices may vary based on the complexity of the survey.

Incorporating Usability Testing for Survey Applications Used by Interviewers

Temika Holland, *U.S. Census Bureau*

The U.S. Census Bureau is making strides towards enhancing the experience of its interviewers using advanced technologies and applications. The intent is to provide a more user friendly, satisfying, cost-effective solution to conduct surveys. One such solution is the Census Operations Mobile Platform for Adaptive Services and Solutions (COMPASS) application used by interviewers during survey procedures. The application includes the questionnaire, case management tools, physical location and security services. Generally, usability testing is conducted in survey research to evaluate the human-computer interface and design of a survey instrument to ensure that it meets the needs, wants, and limitations of potential respondents. This paper provides evidence that usability testing of applications used by interviewers is also important and may reduce measurement error associated with poorly designed applications. Failure to acknowledge the need for usability testing in survey applications used by interviewers can adversely affect data quality. This is especially important if difficulties are encountered in navigating the application, and interviewers are not accurate, efficient, and satisfied with using the new technologies intended to enhance their work. A recent usability evaluation of the COMPASS application revealed important usability considerations for interface design and the added benefits of including experts (e.g., actual Census interviewers) in testing. This paper identifies areas of the COMPASS application that can be improved for increased usability and optimal user experience.

The Effect Usability Testing has on Data Quality: A Design of an Online Diary

Yelena Pens, *The Nielsen Company*

Robin Gentry, *The Nielsen Company*

How do you transform a traditional paper-based instrument that has been in use for 60+ years into a modern web-based application? Nielsen is using an iterative testing process to create such a web-based instrument that could eventually replace paper. Nielsen conducted a series of tests using a probability based address sample to recruit the general population, aged 13 and older, to complete a one week web-based diary of their radio listening. The first web application, or eDiary, pilot test was conducted in 2011. The results of the test identified potential usability issues with the online application such as overlapping time entries and conflicting time issues. Following the field test, a usability study of the online diary application was conducted. Web savvy and non-web savvy participants were invited to use the application and provide feedback on the user experience. The focus of the usability study included content and features of the online diary such as radio buttons, checkboxes, and a slider bar. In winter 2012, an in-home ethnography study was conducted in three markets where participants used scenarios and completed tasks to enter radio listening entries in a diary page. Additionally, expert reviewers were commissioned to review the usability of the overall website from registration to completion of diary keeping. The results from the three studies were consistent and resulted in a re-design of the eDiary application. We will present the results from the eDiary usability study as well as the comparison of the results from the pilot tests. We will show the features of the online diary and the effect usability testing had on data quality such as time precision and duplicate entries. The results show that a simple approach is usually the optimal one.

Challenging Survey Screen Designs on Smartphones

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

Elizabeth Nichols, *U.S. Census Bureau*

With the rise in smartphone and other mobile device usage, survey organizations like the Census Bureau have begun rethinking their Web-survey screen designs. Studies in the market research sector are finding that, on average, 25% to 30% of respondents are choosing to do online surveys using their smartphones, depending on the population being studied. Hence, optimizing for mobile screens is now a priority. Research on the American Community Survey, a 40 minute mandatory survey sponsored by the Census Bureau, has shown that although the percentage of survey responses which come in via a smartphone or other smaller device is still relatively small, the characteristics of those who report via mobile devices is different – they are younger, more likely to be a minority and less educated (Horwitz, 2014). Because the American Community Survey has not been mobile optimized, there are more break-offs when completing the survey on a smartphones. Differential breakoffs by population groups could have measurement error consequences for the ACS. This presentation focuses on results of the smartphone usability testing of the current non-mobile optimized ACS. We conducted 30 interviews with participants in the Washington, DC metropolitan area. They all accessed and completed the current non-optimized ACS on their smartphone or tablet. During the testing, we observed possible errors made while completing the survey - both in terms of changes to the data reported and edits. We then debriefed participants after they complete the survey. This presentation focuses on the major challenges users encounter such as, screen size, the answer interface (aka the fat finger problem), and slow download speeds. Usability results focus on the characteristics of the screens which lead to the greatest potential measurement error during testing.

AAPOR 70th Annual Conference

Young Guns: An Experimental UX Design among 18-20 Year Olds

Paul Schroeder, *Abt SRBI*

Healey Whitsett, *Fors Marsh Group*

Melanie Wilbur, *Abt SRBI*

Brian Griepentrog, *Fors Marsh Group*

The over-representation of young drivers in crashes and road fatalities is a serious public health concern and imposes substantial human, social and economic costs. Given the severity of the problem, NHTSA is conducting the first Youth Traffic Safety Survey (YTSS) to better understand the causes of fatal crash among young drivers 16-20 years old. The YTSS is composed of two questionnaires (version A and version B). Prior to fielding, the survey development included cognitive and usability testing to detect items that were not understood by young respondents as intended and usability issues that impeded them from completing the survey. For the usability testing, three different modes (desktop, mobile device, and paper-based) were tested. One user group began the survey on a desktop and then switched to a mobile device half way through the instrument. Another user group began on a mobile device and switched to a desktop. A third user group completed the survey on paper (half: Version A; half: Version B). Throughout the sessions, we evaluated usability in terms of distribution of responses to specific items, data quality, time to complete survey, length of responses to open-ended questions, placement of clarifying instructions, path taken to complete specific items, and the level of effort required to complete the survey and items. Eye-tracking data was also collected for those who completed the survey on a desktop to evaluate the order in which respondents processed information and worked through the survey. In this presentation, we will discuss and demonstrate the importance of conducting usability testing across modes.

The Web Option: Response Rate Cost and Quality Implications

Trade-Offs in a Survey of HealthPartners Patient Members

Jeanette Y. Ziegenfuss, *HealthPartners Institute for Education and Research*

Juliana Tillema, *HealthPartners Institute for Education and Research*

Kayla Dean, *HealthPartners Institute for Education and Research*

Tom Kottke, *HealthPartners Institute for Education and Research*

HealthPartners has the dual role of insurer and provider of care for approximately 500,000 individuals in several Midwestern States. In order to better understand what data collection strategies work best to ensure high quality self-reported data at a number of price points for this population, we designed an experiment to evaluate pre-notification, mode and incentive structure on outcomes associated with nonresponse bias and costs. Specifically, in a survey assessing the general health and well-being of this population, we randomized 2,550 individuals to a telephone survey with or without pre-notification, a mail survey with or without telephone follow-up and a push-to-web mailing. We also randomized individuals to a no incentive condition, a \$2 non-contingent incentive or a \$10 contingent incentive. For each condition, we compare responders to nonresponders with respect to demographics, health status (i.e., Charlson co-morbidity index) and health care utilization gleaned through combined HealthPartners insurance and medical records. The outcome is a comprehensive set of survey trade-offs that can be used to determine the best strategy for future studies when cost, time and/or generalizability need to be optimized in this and other similar populations.

Web Respondents and Sample Coverage: Is the Gap Between Educational Attainment and Income Closing for the Web Mode? Findings from One National Study

Katie Morrison, *Mathematica Policy Research*

Daniel Friend, *Mathematica Policy Research*

Tiffany Waits, *Mathematica Policy Research*

As access to internet and technology expands across the United States (U.S. Census Bureau 2013), researchers have increasing opportunities to conduct web-based surveys and are reaching a broader population than ever before. Web surveys are an efficient and cost-effective means of collecting data that provide respondents greater convenience and access and allow for easier participation. Despite the seeming ubiquity of internet access, hesitation to use web surveys with certain populations exist because of gaps between demographic groups with regard to access to technological devices and the internet. For example, in 2000, 50 percent of all adults in the United States used the Internet and a large gap existed between high and low income users. Seventy-nine percent of adults with a household income of \$75,000 or higher per year reported using the Internet compared with just 28 percent of all adults making \$30,000 a year or less (Zickuhr and Smith 2012). Yet, in an updated report, Zickuhr and Smith find that the gap between the two extreme income categories has narrowed (97% of adults in the \$75,000+ category reported using the internet in 2011 compared with 62% of adults in the \$30,000 or less category). Research focused on mode differences and sample coverage have found gaps; those with lower socio-economic status (SES) are less likely to complete a web survey than those with higher SES (Israel 2009; Couper 2000). However, preliminary analyses from a multi-modal national study of children and parents appear to indicate that key SES variables (race/ethnicity, educational attainment, and household income) may not be strongly correlated with web response. This paper will examine web survey participation from one study to model and analyze demographic groups and assess the extent to which SES variables may or may not be related to web mode participation.

Web Survey Response Examined from the Perspective of Leverage-Saliency Theory Within a Longitudinal Survey

Yamil Gustavo Nares, *University of Essex*

This research puts forward evidence for potential improvements in web response by taking into consideration the advantages entailed by offering a web mode design and making it salient to random sample members allocated in a mixed-mode design. The framework used to examine data on responses is Leverage-Saliency Theory which argues that multiple survey attributes can be made salient during any survey request. In addition, LST states that each attribute has a potentially different leverage on response for each sample participant. LST allows us to examine web response within a longitudinal survey using a mixed mode design. The basis of this research will be the fifth wave of the (UKHLS) Innovation Panel. One third of households were randomly allocated to a face-to-face survey. For the other two-thirds of the sample a sequential web-CAPI mixed-mode design was employed. The longitudinal aspect of the survey allows for information about mode preferences as well as the provision of respondents' e-mail addresses and a range of other characteristics for respondents and non-respondents to be collected. Results show that non-respondents are likely to become a web respondent if they show a preference for the web and if they have previously supplied their e-mail address. These factors increase the likelihood of achieving an answer online. When being a respondent, the effect of preferring the web and the provision of an e-mail address increases the likelihood of responding online over face-to-face. It should be noted that, controlling for demographic factors, the

AAPOR 70th Annual Conference

interaction is not significant between web mode and supplying an e-mail. However, giving an e-mail is the relevant effect that may have a positive influence on responding online. Interestingly, once a participant becomes a respondent (face- to-face or web), preferring the web and supplying an e-mail address increase likelihood of answering online.

Adding a Web Mode to Phone Surveys: Effectiveness and Cost Implications

Rebecca Lien, *Professional Data Analysts, Inc.*

Harlan Luxenberg, *Professional Data Analysts, Inc.*

Julie Rainey, *Professional Data Analysts, Inc.*

Laura A. Beebe, *Stephenson Cancer Center*

High survey response rates are important when evaluating public health programs, and so are minimizing the associated survey costs. The goal of this study was to assess effectiveness and cost implications of collecting follow-up surveys using a dual mode survey (web and phone) compared to a single mode (phone). Data for the study come from a 2014 evaluation of Oklahoma's Tobacco Helpline. A random selection of participants (n=1,174) were sampled for follow-up and contacted seven months after registering for services. Participants were not randomized into the two study protocols; participants with email addresses received dual mode protocols (n=467) while participants without email addresses received single mode protocols (n=707). Participants registered for services between March 1, 2013 and October 31, 2013. Overall response rate was 47.4% (556 out of 1,174). Effectiveness was assessed by comparing the survey response rates of single mode to dual mode study participants. A multivariate logistic regression model controlled for differences between the two survey groups. Cost implications were calculated by estimating what the cost would have been to conduct two separate surveys: a single mode and a dual mode; the costs of each survey strategy were then compared at different sample sizes. Cost estimates were categorized as fixed (e.g. infrastructure, interviewer training), varying by number of sampled participants (e.g. pre-notification letters), and varying by number of completed surveys (e.g. incentives). After controlling for co-variables, the dual mode survey protocols resulted in a higher survey response rate than the single mode (OR = 1.48, p=.009). However, the overall cost of a study was higher for the dual mode strategy until n=1,100 completed surveys were collected; at which point the dual mode strategy had lower overall costs. The cost analysis provides a helpful framework for assessing cost implications of new survey protocols.

Hope Springs Eternal: Will a Probability Sample of Schools and Principals Respond by Web and Provide Email Addresses?

Cleo D. Redline, *National Center for Education Statistics*

Andrew Zukerberg, *National Center for Education Statistics*

Meta-analyses have shown that response rates for Web surveys tend to be lower on average compared to other modes and that surveys that offer paper and Web modes concurrently also result in lower response rates overall. Even a study of mode preferences showed that offering only a Web survey, while appealing to those who prefer that mode, lowered overall response rates when compared to a mail survey only. Still, the Web is thought of as having advantages over paper that merits its continuous consideration. Recent research in the American Community Survey demonstrated that if not given a choice, people respond by Web. Success stories such as this and the potential advantages of the Web continue to stoke the idea that the Web may yet become a viable survey mode. As people become more adept at using computers they may become more likely to respond by Web. It follows, therefore, that school administrators and principals, who are more likely to have access to and be adept at using

computers, may be earlier adopters of the Web. In this paper, we report on the results of a factorial experiment designed to determine if institutional respondents to three surveys (a teacher listing form, school survey, and principal survey) are indeed more likely to respond by Web than mail. We address whether the advantages attributed to the Web manifest themselves in these surveys and examine respondent compositions by mode, as potential indicators of non-response bias. Finally, administrators were asked in the teacher listing form to provide email addresses for their teachers, so that these teachers could later be contacted by email to participate in a Web survey. We report on the effect that asking for these email addresses has on response rates as well.

Evaluating and Compensating for Non-Response Bias

Identification and Reduction of Nonresponse Bias in Address-Based Sample Surveys

Burton Levine, *RTI*

Dual-frame random digit dialing (RDD) telephone data collection and address-based sampling (ABS) with a mail contact are two commonly used probability survey designs. Data from these different study designs may produce different population estimates for the same outcome due to different sources of survey error. In 2014, the New York Adult Tobacco Survey (NY-ATS) was fielded both as a dual-frame RDD telephone survey and as an ABS survey with mail contact. The two studies produced substantively meaningful different smoking estimates as well as statistically significant differences in other study outcomes. I estimated smoking prevalence for New York block groups by fitting a model using the previous 4 years of NY-ATS data. I found a significant correlation between estimated smoking propensity and response propensity in the ABS sample, but not in the listed landline sample, thereby, demonstrating nonresponse bias in the ABS study. I modified the weighting procedure for the ABS data to use the block group estimated smoking propensity in the nonresponse model. The adjustments based on this model reduced the nonresponse bias in the ABS sample and reduced the difference in the smoking estimates between the ABS and RDD studies. Similar methodology can be applied to studies with outcomes other than smoking.

Examining Differences in Response Propensities and Satisficing Among Medical Providers

Daniel G. Harwell, *American Institutes for Research*

Tandrea Hilliard, *American Institutes for Research*

Alison Huang, *American Institutes for Research*

Melissa Mannon, *American Institutes for Research*

Previous research suggests that the factors that influence the decision to participate in a survey may also affect respondent effort during completion, linking high response propensity with improved data quality (Dahlander, 2012). One specific measure of data quality, satisficing, can occur when respondents shortcut this cognitive process, often due to respondent fatigue or lack of motivation, thus providing suboptimal responses (Krosnick, Narayan, & Smith, 1996; Holbrook, Green, & Krosnick, 2003). One particular population where this is of concerns is with physicians and medical professionals. While physicians and other medical providers are a challenging population to survey in part due to their demanding schedules and the high opportunity cost that completing surveys represents (Sudman, 1985; Flanigan, McFarlane, &

AAPOR 70th Annual Conference

Cook, 2008), they are frequently surveyed. This research utilizes data from a 2014 mixed-mode survey conducted with a stratified random sample of practicing primary care physicians, specialty physicians, physician assistants, and nurse practitioners to determine whether there are differences in survey response propensities and satisficing across these groups. First, we assess survey response (i.e., propensity to respond and timing of response across mail and phone administration modes) among these groups. Additionally, we examine differences in satisficing across provider types by comparing the percentage of respondents who engage in one form of satisficing, which occurs when respondents select the same response option within a series of questions ("straight-lining"). Ultimately, receiving accurate feedback from medical providers is critical to improved health care delivery and quality. Adaptive design approaches may be necessary to increase survey participation and response quality within this key population.

Studying Nonresponse Bias with a Follow-Up Survey of Initial Nonresponders in a National Dual Frame RDD Survey

Paul John Lavrakas, *Independent Consultant*

McKenzie Ballou, *Morris Davis and Company, Inc.*

Deanne W. Swan, *Institute of Museum and Library Services*

Carlos A. Manjarrez, *Institute of Museum and Library Services*

Low response rates and nonresponse bias are major concerns for how well a sample survey represents its target population. However, as Groves (2006) has demonstrated, response rates are not reliable harbingers of nonresponse bias. Furthermore, nonresponse bias must be understood at the individual survey statistics level because within the same survey some measures may suffer from such bias to a non-ignorable extent when others do not. However to date, little has been reported about how such studies have been conducted. The Institute for Museum and Library Services commissioned a nonresponse bias follow-up survey as part of its 2013-2014 national dual-frame RDD survey to measure the use of museums and library services in the U.S. The original survey completed 3,537 interviews with an AAPOR RR3 of 25% for the landline and 10% for the cell phone samples. The follow-up survey of nonresponders from the original survey completed 201 interviews, with an RR3 of 31% for the landline sample and 16% for the cell phone sample. The original sample and the follow-up samples had remarkably similar demographic characteristics with the exception that the follow-up sample had significantly more young adults and fewer old adults. When comparing the weighted results associated with the key behavioral statistics, it was found that for some of the statistics associated with children's use of museums, the original survey appeared to underestimate their prevalence by 5pp or more. In contrast, there were no behaviors which the original survey overestimated by 5pp or more; neither were there any adult behaviors that were over- or underestimated by at least 5pp. Our paper and presentation will report details about (1) how the follow-up survey was conducted and (2) how data from the two surveys were weighted and combined to yield adjusted national estimates of the key behavioral measures.

When a Single Number Won't Do: Methods for Evaluating the Risk of Nonresponse Bias

James Wagner, *University of Michigan*

Surveys are an important source of data for social, behavioral, and health research. However, estimates from survey data may be affected by nonresponse. Unfortunately, an adequate characterization of the risk of nonresponse bias is not possible using only the response rate or any other single number. This paper suggests that in order to evaluate possible nonresponse

AAPOR 70th Annual Conference

biases, researchers will need to adopt a multi-faceted approach. Such an approach would look at nonresponse from a number of different perspectives. This would involve using multiple indicators, including the response rate, R-Indicators, FMI, and model fit statistics. Further, since analysis of missing data requires unverifiable assumptions, a thorough evaluation would also examine the sensitivity of conclusions to these assumptions. Evaluations of the risk of nonresponse bias should carefully examine two dimensions of the problem. The first dimension is the relationship between the complete, auxiliary data (i.e. data from the sampling frame and paradata) and response. The second dimension is an examination of the relationship between the auxiliary data and the survey data from respondents. The latter can include an examination of whether different design features recruit respondents who differ with respect to the survey variables. This talk presents examples of a series of alternative indicators and how they may be used along with sensitivity analyses to develop a constellation of evidence regarding 1) the potential impact of nonresponse on bias and 2) the ability of various design features to mitigate this risk.

Friday, May 15
8:00 a.m. – 9:30 a.m.
Concurrent Session C

Advances in Cognitive Interviewing

Analyzing Cognitive Interviews for Cross-National Studies

Jennifer Kelley, *University of Michigan*

Kristen Cibelli Hibben, *University of Michigan*

Ting Yan, *Westat*

The expanding literature for cognitive interviewing illustrates the importance of conducting cognitive interviews to not only uncover various problems with survey questions (e.g. comprehension or retrieval errors), but to gain a deeper understanding of how survey questions function and if the questions or study constructs are understood as intended by different populations or population subgroups. For questionnaires that are developed for cross-national studies, understanding how the questions constructs perform in different cultural contexts becomes even more crucial if one is aiming for comparability among the countries/cultures. Although there is an emerging literature on cognitive interviewing in the context of cross-cultural studies, the literature on how to analyze the data once collected is sparse. This presentation will review the current literature on cognitive interviewing analysis for cross-national studies, special considerations needed to analyze cross-national data, and a case study drawing on experience analyzing data from cognitive interviewing studies conducted in the United States, Nepal and China.

Tools to Improve Cognitive Interviewing: Q-Bank and Q-Notes

Candace D. Sibley, *National Center for Health Statistics*

Sheba K. Dunston, *National Center for Health Statistics*

Justin Mezetin, *National Center for Health Statistics*

Historically, qualitative research has been critiqued for lacking rigor and systematic assessment in comparison to quantitative research. Common difficulties in qualitative research include conducting team coding and analysis, ensuring transparency of the research, and developing and organizing codes and themes in a systematic way. Studies have discussed the use of software to ease the common challenges of conducting qualitative research and ensure qualitative data are neutral, replicable, credible, and dependable. The Question Design Research Laboratory (QDRL) with the National Center for Health Statistics has developed and manages two original and innovative tools to streamline the qualitative research process for conducting and analyzing cognitive interviews for question evaluation research. Moreover, these tools also assist survey designers in improving questionnaires and gives data users a deeper understanding of survey estimates. The QDRL created Q-Notes, a software program that supports the structured collection and analysis of cognitive interview data. It serves as an audit trail tracing each finding to the original source. The video of the actual interview is embedded within the application so that the findings can truly be traced to the original source. The development of Q-Notes was guided by the theoretical foundations of qualitative research, and specifically customized for cognitive interviewing. Q- Notes also supports replicability or clear paths between the “transcripts, codebook, and analysis notes.” The QDRL has also developed Q-Bank. Q-Bank is a research tool consisting of a database of evaluated questions from Federal surveys and links each question to the scientific report. This presentation will discuss tools that combat the challenges facing the cognitive interviewing field. We will describe how Q-Notes and Q- Bank has met these challenges and how it can be used to assist questions evaluation researchers conduct their work in a streamlined and systematic way. as well as interpret survey questions and understand errors.

Analytic Techniques to Examine Construct Validity in Cognitive Interviewing Studies

Kristen S. Miller, *National Center for Health Statistics*

Cognitive interviewing studies have long been considered a method for determining potential problems in survey questions. However, because cognitive interviewing studies also determine the ways in which respondents interpret questions and apply those questions to their own lives, experiences and perceptions, they also identify the phenomena or sets of phenomena that a variable would measure once the survey data is collected. In this regard, the method is a study of construct validity. Using analytic techniques described in the recently published book, *Cognitive Interviewing Methodology* (Miller, et al, 2014), this paper will illustrate specific methods that can be used to assess construct validity in cognitive interviewing studies. A key theme will be data quality of cognitive interviews and the assessment of that quality when performing analysis and reaching conclusions. To illustrate the methods, the presentation will draw upon a case study that includes 60 cognitive interviews to assess the performance of a set of disability questions intended for international use.

The Value of Pretest Subjects That Don't Represent the Population of Interest

Steven R. Putansu, *George Washington University*

Leading practices in survey research agree that pretest respondents for cognitive interviews should, to the extent possible, represent the population of interest (Willis, 2005). However, due to resource constraints, small target populations that could be tainted by pretesting, and a general reticence among survey practitioners to use cognitive pretesting, this may not be warranted or implemented. With this in mind, this paper seeks to examine similarities and differences in cognitive pretest findings from respondents who do and do not represent the target population. In order to examine the relationship between representing the target population and providing useful cognitive interview responses, this paper compiles and compares cognitive pretests from approximately 50 surveys from a variety of topics. For each survey, 3 to 5 pretests were conducted. For each survey, at least one cognitive interview subject represented the target population and at least one did not. This paper codifies cognitive processes (Encoding, retrieval, decision, and response) that were identified for each respondent and the range of question modifications that resulted from the cognitive interviews. The paper goes on to compare the similarities and differences among response issues and suggested questionnaire improvements, in order to determine the relative value of pretesting subject from outside the target population.

Cells Only and Cell Weighting

Cell Phone Multiplicity: Should Polls Correct for Adults with More than One Cell Phone?

Courtney Kennedy, *Abt SRBI*

Kyley McGeeney, *Pew Research Center*

A sizable proportion of U.S. adults have more than one cell phone, but many telephone surveys assume for weighting purposes that each adult can be reached on only one cell phone. This practice of ignoring cell phone multiplicity simplifies the weighting, avoids increasing the design effect, and avoids adding another question to the instrument. The downside is that failing to adjust for this may result in biased estimates. Adults with multiple cell phones have a higher probability of selection than those with just one cell phone. Multi-cell phone adults would, therefore, be over-represented if no multiplicity adjustment is performed. Computing the adjustment is typically straightforward, with the weighting factor equal to 1 over the number of eligible cell phones (often capped at some value to avoid undue variance in the weights). There is little empirical guidance, however, on two aspects of this issue. First, how exactly should cell phone multiplicity be measured? For example, if an adult has a second phone but never answers it, does that count? Should phones used primarily for work count? Second, does a cell phone multiplicity adjustment demonstrably improve survey estimates or does it simply increase the design effect without reducing bias? To address these concerns, we administered a series of experimental questions in two national dual frame RDD surveys conducted for the Pew Research Center in 2014. In this paper we compare the response distributions to different measures of cell phone multiplicity. We find that as many as 30% of cell respondents report multiple cell phones depending on how the question is asked. We also explore the bias and precision of weighted estimates computed with and without a cell phone multiplicity adjustment. These results will help practitioners make better-informed decisions regarding whether or not to include such an adjustment in their studies.

Testing Dual Frame RDD Surveying of the Japanese General Population

Yasuyuki Saito, *The Asahi Shimbun Opinion Poll Research Center*

Paul J. Lavrakas, *Independent Consultant*

Unlike in Europe and the U.S., regular use of RDD landline surveys in Japan began only in the past decade. Besides concerns about decreasing response rates, recent criticisms of Japanese RDD surveys center on their exclusive use of landline samples. Starting in 2008, Japan's second largest daily newspaper, The Asahi Shimbun, included a question in its very high quality nationwide mail surveys (RR2 ranging from 70%-80%) to measure the proportion of Japanese with home landline service. Since 2010, these surveys included a question about preferences for landline vs. cell phone at home. It was found that Cell Phone Only (CPO) users increased from 15% (2010) to 20% (2014) and Cell Phone Mostly (CPM) users increased from 27% to 32%. Thus, the majority of Japanese now use a cell phone only or mostly. The Asahi Shimbun regards the percentage of CPO/CPM as a critical problem for coverage of Japanese population (compared to using only landline RDD) and therefore will conduct cell phone RDD pilot surveys in late 2014 and early 2015. However, in Japan the cell phone numbering system includes no regional information, unlike the landline system. So it is impossible to conduct a regional cell phone survey in Japan without very costly screening. Thus, for the inaugural pilot surveys nationwide cell phone RDD surveys will be conducted during the same period as nationwide landline RDD surveys are conducted. Our paper will describe the method for creating RDD telephone numbers for the Japanese cell frame, how to best sample those numbers, CATI data collection methods for Japanese cell phone users, and comparisons between the RDD landline and RDD cell phone surveys' key statistics and between dual users of cell and landline phone versus cell phone only users, including for respondents' characteristics and opinions about current political issues.

The Effects of Overlapping RDD Sampling on the BRFSS

Carol A. Pierannunzi, *Centers for Disease Control and Prevention*

Pranesh Chowdhury, *Centers for Disease Control and Prevention*

Machell Town, *Centers for Disease Control and Prevention*

The Behavioral Risk Factor Surveillance System (BRFSS) is a continuously collected system of telephone surveys of health practices and health risk behaviors in each of the U.S. states, DC and several U.S. territories. Although the BRFSS was traditionally a landline sample, cell phone interviewers were added to the publicly released BRFSS dataset beginning in 2011, using a dual frame. In 2013, interviews with cell phone respondents were included if the respondent reported that he/she mostly used a cell phone for incoming calls. In 2014, a fully overlapping sample was implemented. This research examines the impact of the changes in sampling due to the move from a single sample, to dual frame, to partially overlapping and finally to fully overlapping. Analyses use the BRFSS 2011-2014 datasets. Demographic comparisons are examined for each of the sampling methods. Differences in health outcomes are also compared across types of samples. Finally an assessment of the efficiency of the new methods is examined and comments are made about the future of moving to an all cell phone sample. Efficiency is measured by the number of completed interviews per hour and the use and size of the sample. Results indicate that the overlapping sample improves unweighted demographic representation and provides for more efficient use of resources. The impact of the overlapping sample on health outcomes is less distinct, potentially due to the effects of raking weighting.

Transition from Landline-Cell to Cell Frame Design: Surveys of Consumers

Charley Jiang, *Institute for Social Research University of Michigan*

James M. Lepkowski, *Institute for Social Research University of Michigan*

Tuba Suzer-Gurtekin, *Institute for Social Research University of Michigan*

Michael Sadowsky, *Institute for Social Research University of Michigan*

Richard Curtin, *Institute for Social Research University of Michigan*

Rebecca McBee, *Institute for Social Research University of Michigan*

Dan Zahs, *Institute for Social Research University of Michigan*

The University of Michigan's Surveys of Consumers is moving from a dual-frame landline and cell telephone sample design to a single-frame cell telephone design in 2015. The change is being made in part to eliminate complexity in dual-frame weighting, and in part to take advantage of demographic distributions that provide more completed interviews with 18-34 year olds. In addition, 93% all U.S. persons 18 years and older currently have cell phones. The transition to cell frame sampling will provide comparable coverage to an era more than a decade ago when landline-frame sampling had similar coverage rates and was thought satisfactory. The transition will change the Surveys response rates. Current dual-frame rates are compared to projected cell-frame rates, showing a continued decrease in response rates for the Surveys of Consumers. The transition will also change the demographic composition of the responding sample, and dual-frame unweighted respondent composition is compared to projected cell-frame respondent composition. The larger proportions of male interviews and interviews with 18-34 year olds, as well as other differences, are shown. Finally, the current cell-frame sample weighting system is a component of the current dual-frame weighting system. Eliminating the landline-frame sample will produce a new weighting system for cell-frame only sample. The current weighted demographic and consumer attitudes estimates are compared to projected cell-frame only sample corresponding weighted estimates to show the potential impact of changes to key estimates during and after the transition to cell-frame only sample design. Charley Jiang, James M. Lepkowski, Tuba Suzer-Gurtekin, Michael Sadowsky, Richard Curtin, Rebecca McBee, & Dan Zahs Institute for Social Research University of Michigan

Health Surveys and Self Reports: Exploring Reliability and Validity

Examining Measurement Error in Self-Reports of Receiving Financial Assistance under the ACA

Daniel G. Harwell, *American Institutes for Research (AIR)*

Survey researchers attempting to measure the impact of the Affordable Care Act (ACA) face a critical challenge given the complex and technical nature of reforming health insurance, which then must be communicated to respondents through survey questions utilizing non-technical terminology. One of the most complex aspects of the ACA involves the specific tools utilized to provide financial assistance for individuals to purchase health insurance: the Advance Premium Tax Credit (APTC) and cost-sharing reductions (CSR). Both of these mechanisms have been in place since January 1, 2014, yet many consumers may not understand how or why the cost of their health insurance has been reduced, or may not realize that that the cost of their health insurance has been reduced. This research examines the error rate of respondent self-reports of receiving assistance in paying for their health insurance, including both the APTC and CSR. More specifically, we identify discrepancies between self-reported survey data from the 2014 field test of the Health Insurance Marketplace Experience Survey to administrative records from

AAPOR 70th Annual Conference

the Centers for Medicare & Medicaid Services (CMS). Preliminary research shows that there are differences between the self-reports and administrative data, with respondents under-reporting. In addition to measuring the rate of under reporting, this research will utilize a logistic regression to identify populations that are more likely to have discrepancies between their self-report and administrative data. Finally, we examine the differences and similarities of the findings from this analysis with previous research examining underreporting of Medicaid enrollment in other government surveys in order to better understand the sources of this error.

Reliability and Validity of Self-Reported Health Status: Two Measures of Self-Reported General Health Status in the National Health and Nutrition Examination Survey (NHANES)

Hee-Choon Shin, *NCHS*

Jibum Kim, *Sungkyunkwan University*

In general, it is known that self-reported general health status is a good proxy measure for actual health and mortality. The National Health and Nutrition Examination Survey (NHANES), a periodic survey conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), is a program of studies designed to assess the health and nutritional status of adults and children in the United States. NHANES is a nationally representative survey of the resident civilian non-institutionalized U.S. population. The survey is unique in that it combines interviews and physical examinations. It consists of questionnaires administered in the home, followed by a standardized physical examination in a specially equipped mobile examination center (MEC). The examination includes physical measurements such as blood pressure, a dental examination, and the collection of blood and urine specimens for laboratory testing. In NHANES surveys, self-reported general health is measured twice – first, in a home interview, and then in an MEC interview. The MEC is configured from four connecting tractor trailers. Health professionals conduct the health exams. The survey includes two exam teams. There are 16 individuals on each traveling team: 1 MEC manager, 1 MEC coordinator, 1 licensed physician, 3 medical technologists, 4 health technologists, 2 registered dental hygienists, 2 MEC interviewers, 2 dietary interviewers, and 1 phlebotomist. In addition, local assistants are recruited, trained, and employed at each stand to assist the exam staff. Our objective is to study the reliability between, and validity of, the two measures by analyzing the relationships between them and measured health indicators (including blood pressure and cholesterol levels). The 2011-2012 NHANES public use file will be used for data analyses. A preliminary study shows that there is a substantial discrepancy between the two self-reported general health indicators.

Physical Activity: Measurement and Self-Reports

Arie Kapteyn, *USC*

Tania Gutsche, *USC*

Bas Weerman, *USC*

Variations in physical activity across socio-economic and demographic groups are of great importance for understanding differences in health behaviors and their consequences. There is strong evidence however that self-reports of physical activity are subject to systematic response scale differences across these same demographic and socio-economic groups. To address these concerns, we are conducting an accelerometer study among a population representative sample of approximately 750 adults who are asked to wear an accelerometer for a period of 8 days. In addition we administer a questionnaire asking for various measures of self-reported physical activity, while moreover respondents are shown a number of anchoring vignettes

AAPOR 70th Annual Conference

describing hypothetical persons exhibiting different physical activity levels. This set-up allows us to calibrate self-reports by comparison with the data obtained from the accelerometers; the anchoring vignettes can potentially be used to correct for differences in response scales across respondents. The physical measurements allow us to test how accurate such corrections are. We have earlier collected similar data in the Netherlands. In the paper we will compare physical activity and self-reports both in the U.S. and the Netherlands

Four Well-Being Dimensions Predict the Obesity and Disease Burden

Diana Liu, *Gallup Organization*

Andrew Dugan, *Gallup Organization*

Using Gallup's Daily Poll data for year 2014, this research will identify the inner relationships between five well-being dimensions including Physical, Purpose, Financial, Social and Community. This presentation will show that four of these dimensions of well-being can all predict the obesity and disease burden, a metric which is based off of self-reported responses to questions regarding physical health. Multiple regression analysis is used to analyze the inner relationships between the five well-being dimensions. Logistic regression is applied to study the associations between obesity/disease burden and four well-being dimensions. Also, an annual cost of disease burden is estimated. People with higher well-being scores in the areas of social, community, financial and purpose are less likely to be obese, after controlling for disease burden and demographic differences. Additionally, people with higher well-being scores in the areas of social, community, financial and purpose are more likely to have lower disease burden, after controlling for obesity and demographic differences.

Vietnam Population Health Survey

Judith M. Tanur, *Stony Brook University*

Atlantic Philanthropies (AP) has made several large grants over the last decade to help improve healthcare in Vietnam, especially for the underserved. They have built new commune health centers, trained doctors as family practitioners, donated equipment, and incorporated programs to encourage safe motherhood and healthy infancy. AP has worked in conjunction with provincial governments and has created a method that they think is exportable to other third-world countries. To see if this effort is successful, AP has made several grants to the Vietnam Program at the Social Science Research Council (SSRC) for an evaluation. SSRC works with 4 American-trained Vietnamese PI's and an international advisory committee to carry out a longitudinal survey (n~ 3600 households in 12 communes), administered early in the interventions and again later and also to follow up in each round with in-depth interviews with some 360 households, exit interviews of patients or their care-givers as they depart commune health centers and other health care providers, with surveys of the equipment, staffing, and procedures of these same providers, and with participant observation in each of the study communes. As a member of the international advisory committee, I helped design the survey, worked on the analysis of the first round of data and am at work on the analysis of the second round. My presentation will describe some of the challenges of this sort of international work and present some of the early results of the analysis, including a finding that the presence of a doctor trained in family medicine is associated with a statistically significant change in self-perceived health status.

The “Good Enough” Respondent: Using Sufficiently Knowledgeable Adults as Respondents in a Survey about the Health of Young Children

Alicia M. Frasier, *NORC at the University of Chicago*

Jennifer Vanicek, *NORC at the University of Chicago*

Andrea Mayfield, *NORC at the University of Chicago*

Ying Li, *NORC at the University of Chicago*

Ned English, *NORC at the University of Chicago*

Joelle Greene, *Harder+Company Community Research*

Melinda Leidy, *First 5 LA*

In surveys about children’s health and well-being, it can be challenging to efficiently identify and interview an appropriate adult respondent. Researchers often decide whether any household member who knows the child can answer questions or whether a more knowledgeable respondent is necessary. The First 5 LA Family Survey was developed to provide representative data about key indicators of well-being for children ages zero through five across 14 communities in LA County, CA. The study employed a multi-mode address-based sample (ABS) design with the aim of efficiently interviewing households with young children. While surveys about children often attempt to interview the adult who knows the most about the child (Most Knowledgeable Respondent, or MKR), research conducted by Eisenhower et al. (2012) suggests that alternate adults may provide equally knowledgeable responses with lower screening and administration costs and less survey error. Building on these findings, the Family Survey was designed to identify and interview a Sufficiently Knowledgeable Respondent (SKR) for each of up to four children in the household. An SKR was defined as any resident adult who reported they could answer questions regarding the selected child. In addition, the Family Survey included questions used by Eisenhower et al. to determine whether each SKR was also the MKR, or whether the household also contained an MKR. This paper examines the effect of interviewing SKR vs. MKR in this health survey about young children in LA County. We compare certain indicators to explore possible differences among these groups. We attempt to determine whether any changes are due to the respondent or the nature of the questionnaire. This paper attempts to further validate use of SKRs as respondents in surveys about young children’s health and well-being. Findings have potential implications for decreased administration costs and screening time in other multi-mode surveys requiring proxy respondents.

Measuring Race and Ethnicity

Mixed-Race Americans: Who Counts and How Do We Count Them?

Eileen Patten, *Pew Research Center*

Juliana Horowitz, *Pew Research Center*

Rich Morin, *Pew Research Center*

Scott Keeter, *Pew Research Center*

The mixed-race population in the United States is almost certainly growing, yet there is no scholarly consensus on who is considered mixed race and how many mixed-race Americans there are. To explore how mixed-race Americans formulate their racial identity and how different approaches to measurement produce widely varying estimates of the mixed-race population, we conducted a series of experiments with a nationally representative sample of adults in Pew Research Center’s American Trends Panel. The resulting estimates of the mixed-race adult

AAPOR 70th Annual Conference

population vary from a low of 3.7% to a high of nearly 20%. The different forms analyzed are: (1) standard telephone approach allowing respondents to specify more than one race (3.7%), (2) Census' Alternative Question Experiment (AQE) (4.8%), (3) AQE plus single-race respondents who say they have a parent (13%) or grandparent (19.8% combined) of a different race, (4) asking respondents to allocate 10 points across racial categories (12.9% allocated 1 or more point to two or more races) and (5) an attitudinal question asking respondents explicitly whether they consider themselves mixed race (12%). The variability of incidence indicates the imprecise nature of defining a mixed-race population. The panel will allow us to look at responses across the different measure of race to see who selects in or out depending on the question asked—for example, our preliminary analysis shows that 6% of single-race respondents to the AQE consider themselves "mixed race" and 30% of multiple-race respondents do not. Furthermore, measurement is affected by the inclusion or exclusion of Hispanic as a racial category (the current figures, for the most part, do not include those who select Hispanic and only one other race). The study also will report the results of embedded experiments testing mode of interview effects (telephone vs. online), as well as order effect for the attitudinal question.

Comparing Methods and Findings Across Focus Group Projects with Alaska Natives to Improve the Race Question for the 2020 Census

Laurie Schwede, *Center for Survey Measurement, U.S. Census Bureau*

Rodney L. Terry, *Center for Survey Measurement, U.S. Census Bureau*

Leticia Fernandez, *Center for Survey Measurement, U.S. Census Bureau*

Alaska Natives and American Indians (AIAN) are classified as one racial group in U.S. statistics, but each subpopulation has great internal diversity. There are more than 120 distinct Alaska Native groups. What is the better way to word the AIAN census race question and tribal instructions to improve data quality for this hard-to-count and hard-to-classify population? And what focus group methods are better to address this question? This paper compares and contrasts both methods and findings from two successive linked focus group projects with Alaska Natives on 2020 Census race question alternatives. In the first project--the 2010 Census Alternative Questionnaire Experiment Focus Group Evaluation--five focus groups were conducted with Alaska Natives in four Alaska sites, with a relatively fixed focus group approach to get answers to standardized questions. Three race versions were included. The second (followup) study--the 2014 AIAN Race Research Project--built on those earlier results, comparing responses to a revised version from the previous project, the 2010 Census control form and four new versions, in two focus groups. In contrast to the earlier site-specific recruiting methods, researchers worked with the Alaska Federation of Natives (AFN) at their annual Anchorage convention, drawing participants from all over Alaska. Participants recruited from multiple ethnicities and areas were less likely to be acquainted than before. A new moderator guide was developed to stimulate greater interaction and explore key concepts. This paper also documents how this followup focus group research with a hard-to-classify population can both check previous qualitative findings and also permit more in-depth exploration of racial classification issues. The paper closes with both substantive and methodological recommendations on 1) improving the race question instructions for Alaska Natives for later 2020 Census research and 2) criteria to use in designing focus groups methods for small, diverse, hard-to-reach populations, such as Alaska Natives.

Race and Ethnicity: Cognitive Testing of a Combined Item for the American Community Survey

Darby M. Steiger, *Westat*

Martha Stapleton, *Westat*

Hyon Shin, *U.S. Census Bureau*

Merays Rios, *U.S. Census Bureau*

Measuring race and ethnicity has been at the forefront of research in the decennial census since the 1800s (Gauthier, 2002; Okazaki and Sue, 1995). Not surprisingly, finding a classification scheme that accurately reflects individuals' racial and ethnic identity have proven challenging. Given the complexity of these constructs, the Census Bureau has tested many variations of the race/ethnicity questions over the years in an attempt to improve measurement accuracy. In addition, Terry and Ford (2013) conducted a test with monolingual Spanish-speaking respondents to compare two versions of the race and ethnicity items—in one version the question is combined and in the other version the items are asked in the more familiar two-item format. The current project extends the work of Terry and Ford, where some respondents received race and Hispanic origin question items in separate questions and others were asked about these constructs in a combined item. The combined item also included a new Middle Eastern or North African category. We will present findings from three rounds of cognitive testing with approximately 70 English-speaking and 60 monolingual Spanish-speaking respondents across the United States and in Puerto Rico. In the first round, we tested both the one-item and two-item versions. In the second and third rounds, we tested variations of the one-item version, which also included a neutrally-worded question stem ("Which categories describe Person 1? Mark all boxes that apply AND print details in the spaces below.") We explored how well the questions fit with the way respondents think about their origin and race; preference for one version over the other; and reactions of particular groups, such as those with multi-racial backgrounds and foreign-born respondents with U.S. born children.

Comparing Methods and Findings of 2020 Census Research on American Indians' Responses to Alternative Race Question Instructions

Rodney L. Terry, *U.S. Census Bureau*

Leticia E. Fernandez, *U.S. Census Bureau*

Laurie K. Schwede, *U.S. Census Bureau*

Among the race groups recognized by the U.S. Office of Management and Budget (OMB), the American Indian or Alaska Native (AIAN) category poses a unique situation to respondents because it includes the provision of maintaining tribal affiliation or community attachment. The 2020 Census race question has a checkbox for the AIAN category, as well as an instruction to print the name of the "enrolled or principal tribe" in a set of write-in boxes. Previous qualitative research and consultations with AIAN stakeholders have shown that many American Indians find the current tribe instruction confusing or not inclusive of their tribal affiliation or community attachment. To address these challenges, the Census Bureau has conducted 2020 Census research on American Indians' responses to race question instructions. We first discuss methods and findings from seven focus groups conducted as part of the 2010 Alternative Questionnaire Experiment (AQE), which identified question format strategies for asking about race and Hispanic origin in the 2020 Census. We then compare those methods and findings with those from four focus groups conducted to inform which of six alternative race question instructions were chosen for a split-ballot 2015 National Content and Self-Response Test. These six alternative tribe instructions differ in the terms used to describe group attachment, for

AAPOR 70th Annual Conference

example, "enrolled or principal tribe," "tribe or village," and "specific origin." Methodological differences discussed include: (a) redesigning the moderator guide to increase free-flowing discussion, (b) using respondent recruitment strategies that also accounted for whether respondents were from federal-, state-, and non-recognized tribes, and (c) using focus group moderators who were not American Indian. We then discuss how these methodological differences may have impacted focus group outcomes, including the amount of respondent feedback and respondents' perceptions of the race question instructions. Finally, we discuss the wider survey implications of both studies.

Comparison of Focus Group Methods and Findings Among Mexican, Central and South American Individuals Answering Questions About Race and Identity

Leticia Fernandez, *U.S. Census Bureau*

Laurel Schwede, *U.S. Census Bureau*

Rodney Terry, *U.S. Census Bureau*

The 2010 Census found robust increases in the size and diversity of the American Indian or Alaska Native (AIAN) population. In particular, the AIAN population increased from 4.1 million to 5.2 million since 2000, with one out of five identifying also as Hispanic (Norris et al. 2012, Humes et al. 2011). The self-identification of Hispanics as AIAN may be negatively impacted by two issues which have not been fully explored. First, the write-in instructions in the Census forms ask for "enrolled or principal tribe," which may discourage reporting among indigenous Hispanics who are not likely to be enrolled, and who may identify with a village or language rather than with a tribe. Second, earlier research suggests that the racial identification of Hispanics who identify as AIAN may be situational and fluid, although the reasons are not fully understood. In this paper we compare methods and findings from two sets of focus groups conducted in Spanish among indigenous Hispanic immigrants to assess barriers to reporting as AIAN. The 2010 Race and Hispanic Origin Alternative Questionnaire Experiment (AQE) included focus groups that collected information about alternative formats for the race and Hispanic origin questions, among them one that combined these questions into a single item. Three of these focus groups were conducted among indigenous Central and South American participants. More recently, as part of the 2020 Census research program, a follow-up study used the same combined format to further probe Hispanic participants in two focus groups about six alternative versions of write-in instructions for the AIAN category. In this paper we discuss the barriers that were mentioned by participants in both sets of focus groups, as well as the differences that emerged across various dimensions. We conclude by discussing the methodological implications and potential issues associated with comparing findings across focus groups.

Using Paradata During Data Collection and in Data Analysis: New Metrics to Address Perennial Problems

Using Paradata During Data Collection and in Data Analysis: New Metrics to Address Perennial Problems

Julia F. Coombs, *U.S. Census Bureau*

Rachael Walsh, *U.S. Census Bureau*

Survey researchers have increasingly acknowledged paradata's usefulness in the monitoring of data collection and in the analysis of error. Kreuter's comprehensive book, *Improving Surveys*

AAPOR 70th Annual Conference

with Paradata: Analytic Uses of Process Information (2013), demonstrates the practicality of collecting and using data about the data collection process. The variety and complexity of paradata available to survey practitioners invite innovative implementations that improve the data collection and analysis process. This panel will present four new uses of paradata that address challenges faced by survey researchers; Gabriele Durant will then unify the presentations through discussion. The first paper will describe the properties and limitations of the scaled evenness of finding attempts (SEFA), a measure of the spread of contact attempts made to establish initial contact with a case across time that is interpretable at the interviewer level. The second paper will use doorstep concerns recorded in the U.S. Census Bureau's Contact History Instrument to examine trade-offs between error and costs. The third paper will explore the relationship between interview complexity and data quality. An index of complexity will be created through confirmatory factor analysis of collected paradata, and it will be compared to error rates in ATUS and SIPP. The final paper will use call history logs and a rich survey frame in the Commercial Buildings Energy Consumption Survey to determine if buildings of different sizes and functions should be contacted in different ways.

Locating Respondents: Designing a Scale to Rate Interviewer Effort

Julia F. Coombs, *U.S. Census Bureau*

Rachael Walsh, *U.S. Census Bureau*

Despite the regulated nature of survey research, the scheduling and management of contact attempts for face-to-face surveys are usually decided by field interviewers. Interviewers are often given a workload and are asked to complete all interviews at their discretion. While some survey organizations offer general guidelines for making contact and achieving cooperation (Stroop et al. 2010), experiments that assign specific contact times to cases have not been successful (Wagner et al. 2012). Interviewers seem to be skilled at interpreting the day-to-day dynamics of fieldwork, and, for now, they remain the best resource for completing face-to-face interviews. Instead of telling interviewers when they should make contact attempts, we focused on quantifying interviewer behavior. We developed a measure of the variation of the timing of contact attempts made when an interviewer is establishing first contact with a household. The scaled evenness of finding attempts (SEFA) quantifies the evenness of an interviewer's attempts to contact a case across time. A score of one indicates evenly spread contact attempts, while a score of zero means that contact attempts were not distributed across time windows. Since the respondent can influence the timing of contact attempts, only the attempts made up to and including the first contact with a respondent are used to calculate SEFA. An interviewer's average SEFA across all cases thus quantifies the interviewer's approach to timing contact attempts. SEFA was first introduced as the contact attempt diversity index (CADI) and was used to model the number of contact attempts to resolve a case (Coombs and Walsh 2014). Additional work included SEFA when modeling the outcome of a case (Walsh and Coombs 2014). This paper focuses solely on SEFA, studies its properties and limitations, and develops the index as a potential indicator of interviewer effort.

Use of Doorstep Concerns to Examine Trade-Offs Between Error and Costs

Ting Yan, *Westat*

Shirley Tsai, *Bureau of Labor Statistics*

Doorstep concerns – one type of paradata – capture the interactions between interviewers and potential survey respondents during the survey introduction and reveal the concerns sampled members have expressed about the survey request and also their reasons for refusing the survey request when refusal occurs. We've created two parsimonious measures that retain the

AAPOR 70th Annual Conference

interrelationships inherent in the doorstep concerns data – the Perceived Concerns Index (through principal component analysis) and the Reluctance Class (via latent class analysis) – and have demonstrated that both measures are effective in characterizing and assessing the level of reluctance exhibited by potential survey respondents. Taking advantage of data from the Consumer Expenditure Interview Survey (CE), this paper explores the trade-off between the cost in pursuing respondents with different levels of reluctance (as measured through doorstep concerns) and the potential error in the resultant expenditure estimates. Specifically, we examine several scenarios in which respondents with different levels of reluctance are excluded. Then we will compare statistical estimates of expenditures derived in each of the scenarios to estimates derived from including all cases. We will also explore the impact of excluding respondents with different levels of reluctance on multivariate relationships. Through these sets of analyses, this paper demonstrates how doorstep concerns data can be used effectively to monitor cost and error in the later rounds of data collection in a longitudinal survey setting.

The Use of Paradata to Evaluate Interview Complexity and Data Quality (in Calendar and Time Diary Surveys)

Ana Lucía Córdova Cazar, *University of Nebraska, Lincoln*

Robert F. Belli, *Survey Research and Methodology Program, University of Nebraska-Lincoln*

To improve data quality, researchers evaluate, control, and compensate for errors that arise from different sources. Furthermore, error prevention and quality improvement can only be achieved by carefully considering the processes that generate the data (Biemer and Lyberg, 2003). Thus, researchers are increasingly using paradata as an additional tool that assesses the impact of data generation on data quality (Kreuter, 2013). Although informative, paradata usually involve very large and complex datasets, as the units of analysis (keystrokes, mouse-clicks) are nested within higher levels (respondents, interviewers). Such complex structure needs to be correctly considered for paradata to validly inform the assessment of both the survey process and the resulting data quality. Building on previous research, in this paper I will create an interview complexity index using substantive answers and paradata from two surveys: the 2010 American Time Use Survey (ATUS) and the Survey of Income and Program Participation (SIPP). Specifically, for each survey I will conduct a confirmatory factor analysis to test whether the covariance between measures such as interview duration, changes in answers, back-ups, number of prompts that appeared during the interview, total number of entries, and total activities/programs reported can be accounted by an underlying factor of interview complexity. By estimating a structural equation model, I then explore the relationship between this interview complexity index and data quality in each of the surveys. Data quality in the ATUS will be assessed by looking at two types of error (ATUS error 1: providing insufficient detail, and ATUS error 2: having memory gaps) and in the SIPP by looking at the correspondence between survey responses and administrative records.

A Tailored Respondent Approach: Developing a Response Propensity Model for a National Building Survey

Carolyn Hronis, *U.S. Energy Information Administration*

Katie Lewis, *U.S. Energy Information Administration*

The Commercial Buildings Energy Consumption Survey (CBECS) is the only nationally-representative survey of the U.S. commercial building stock and its corresponding energy use. Data are collected during a computer-assisted interview, with an average of 9 contact attempts

per complete interview. The unit of analysis is the building, from small, simple establishments about the size of a house to large, complex buildings, such as hospitals and high rises. Because of the diversity in building characteristics and variable availability of knowledgeable respondents, we would like to look at how to tailor our respondent approach for future CBECS rounds in order to maximize response rates, while reducing burden and cost. Using data collected from the 2012 CBECS, we conduct analysis to develop a response propensity model for use in future survey rounds. We have access to a variety of paradata, including call history logs and survey frame files. Unlike household or business surveys, EIA constructs the CBECS frame using both traditional area listing and targeted lists for special populations, such as federal buildings and airports. As a result, the frame has rich data, including estimates of square footage and building activity, both of which are predictive of building energy use, our variable of interest.

Evaluating "Response Rates" for Web Surveys

Do Response Rates Matter in Online Panels? Comparing the Representativeness at Different Levels of Cumulative Response Rates in a Probability Based Online Panel During Several Consecutive Waves

Johan Martinsson, *University of Gothenburg*

Karolina Riedel, *University of Gothenburg*

The last decade several studies have reported that the link between nonresponse rates and nonresponse bias is not as strong as commonly believed. The link between response rates and the quality of a survey seem to be weak. Thus, response rates can only serve as a quite crude indicator of data quality, if it contains any meaningful information at all. Further, the rise of online surveys and the use of online panels have led to many surveys being unable to report proper response rates due to the use of opt-in panels. For such surveys, it is doubtful if the participation rates are at all a meaningful indicator of data quality. And for probability based panels, cumulative response rates (relating to the initial population sample used to recruit panelists), are often very low, or not reported. This study examines the relationship between different levels of response rates within the same survey. This is done through a large scale recruitment effort where those invited to the panel were randomly assigned different levels of recruitment effort and follow up contacts. The recruitment rates vary between 5 and 24 percent. This study tracks these respondents during seven waves (two years) of data collection and examines how the cumulative response rates develop over time and if the groups with higher recruitment rates, and ensuing higher cumulative response rates, are more representative of the general population. We examine representativeness in terms of demography by comparing to official register based statistics, and in terms of political attitudes by using a benchmark high quality survey, and finally to election results.

Boosting Probability-Based Web Survey Response Rates via Nonresponse Follow-Up

Angela Fontes, *NORC at the University of Chicago*

Kean Chew, *NORC at the University of Chicago*

Paul Lavrakas, *NORC at the University of Chicago*

Boosting Probability-Based Web Survey Response Rates via Non-Response Follow-up By
Angela Fontes, Kean Chew, and Paul Lavrakas Web surveys using probability samples have

AAPOR 70th Annual Conference

the advantage of known probabilities of selection for each sample unit and typically produce more representative samples than non-probability sample sources; however, such surveys typically have low AAPOR responses rates for a variety of reasons. This paper describes a case study for boosting the response rate for web surveys sourced from probability samples. The response-rate boosting methodology involves a few components: conducting the initial web survey with a probability sample; drawing a random sample of the non-responders (excluding ineligible sample units) for nonresponse follow-up; and conducting the non-response follow-up study using professional field staff recruiting the sample to participate in the web survey. In addition to describing the methodology, the paper will present perspective on the appropriate calculation of a weighted response rate and standard errors from the combined sample (sourced from the initial survey and the non-response follow-up study). The paper will provide statistical evidence on impact on the sample representativeness and statistical findings as a result of including the interviews obtained from the non-response follow-up study.

Testing Email Invitations in a Non-Probability Panel

Elizabeth M. Nichols, *U.S. Census Bureau*

Ryan W. King, *U.S. Census Bureau*

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

As we move more and more to Web-based data collection and away from paper surveys, it makes sense to notify the sample via electronic means as well. Email and text messaging are replacing mail and phone calls in our day-to-day communication. However, the optimal email or text message to motivate a survey respondent is still quite unknown. Who should the sender be? What should the subject line be? When should I send the email? Do the same characteristics of paper invitations - the sender, the topic still hold, or are there newer, more important characteristics of the email or text that matter more? This paper discusses recent findings regarding email invitations for survey participation. Since March 2013, the U.S. Census Bureau has offered the ability for the public to sign up to participate in online research studies. The sign up is through the Census Bureau website and there is no monetary incentive associated with it. Since its inception, approximately 1,000 people have signed up each month to participate. To learn about optimal email invitations to generate a survey response, six different tests occurred using this Census-affinity nonprobability panel. Email characteristics manipulated included the email subject line, email content, email send time, and email format. Results indicated that a text-based email format outperformed a graphical email format; a number in the subject line "10-minute U.S. Census Survey to Help Your Community" generated more email opens than other email subject lines; and changing the survey burden from 5-minutes to 10-minutes decreased click-throughs to the survey. This paper focuses on the findings from the email testing using the nonprobability panel and future plans to replicate these findings with other probability-based panels.

Does Providing an Email Address in an Initial Contact Study Indicate Respondents Will be Cooperators in a Subsequent Online Panel Study?

TraShawna Boals, *Experian Marketing Services*

Kevin Jean, *Experian Marketing Services*

Research has suggested that using email as a communication channel can increase response rates in online studies. However, does providing an email address also suggest that these individuals are more likely to be cooperators in future online or offline research? This study examines how the cooperation rate for a future study differs between individuals who are willing to provide an email address in the initial phase and those who are not. A two-phase

AAPOR 70th Annual Conference

methodology was used, and all respondents who completed an initial recruitment phase survey later received a mail invitation to a future online panel study. In the first survey phase, respondents were asked to provide an email address and about 30% of respondents complied. Half (approx. 2,500 respondents) of these respondents were then sent an email invitation to participate in the second phase of the survey, an online study, in addition to the mail invitation that was sent. Results indicate that respondents who provide an email address in the initial contact study are more cooperative in joining a subsequent online panel than those who do not provide an email address. Interestingly, for the subsequent study, those respondents who provided an email address and were never sent an email invitation responded to the mail invite at about the same rate as those who received an email invitation.

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

Kristen Cibelli Hibben, *University of Michigan*

Frederick Conrad, *University of Michigan*

To the extent that the inaccuracy in survey responses is due to insufficient effort by respondents, it might help to directly ask respondents to try harder and elicit an explicit agreement from them to do so. The rationale for this technique is that agreeing or stating one's intention to behave in a certain way commits a person to carry out the terms of the agreement. Charles Cannel and his associates pioneered the technique in the 1970s in face-to-face health-related interviews and the results were promising: respondents who signed a commitment statement reported more health events and their reports were higher quality than did respondents who were not asked to make this commitment. A high impact domain in which the quality of survey responses has been called into question – and which therefore might benefit from more committed respondents is labor force economics, in particular, income and the receipt of unemployment benefits (Moore, Stinson, and Welniak, 1999; Meyer, Mok, and Sullivan, 2009). The current study measures the impact of respondent commitment to provide complete, accurate, and honest answers in this domain in an online survey, i.e., without an interviewer present. The data come from a survey conducted by the Institute for Labor Market and Occupational Research (Institut für Arbeitsmarkt und Berufsforschung (IAB)) in Germany. Respondents will be asked to check a box indicating their commitment. Response accuracy to several employment and income related questions will be validated using IAB's employment records– previous studies evaluating commitment have only used indirect measures of accuracy. The study also examines the effect of respondent commitment on socially desirable reporting, satisficing, respondent-reported record check, consent to link their records and responses, and to be contacted for a follow-up survey. If successful, commitment in online surveys is a low cost and low-tech approach to improving data quality.

Predictors of Completion Rates in Online Surveys

Noble Kuriakose, *SurveyMonkey*

Jon Cohen, *SurveyMonkey*

Sarah Cho, *SurveyMonkey*

Mingnan Liu, *SurveyMonkey*

The use of online surveys is now ubiquitous, but there continues to be gaps in survey research about best practices for web questionnaire design. Further, established best practices for web surveys have often been based on research conducted among small samples. In this paper, we aim to improve upon these best practices by analyzing surveys created and conducted on the SurveyMonkey platform, which receives millions of responses to user-created surveys daily. Using this massive dataset, we will glean insights about optimizing the length and types of

AAPOR 70th Annual Conference

questions in an online survey. We examine the effects of several factors on survey completion rates, including survey length, types of questions deployed (such as open-ends and matrix-style questions), number of pages, number of answer choices, and other related measures.

Factors Related to Survey Participation

The Intersection of Sampling and Nonresponse: Does Repeated Sampling of Some Individuals Affect Nonresponse Bias?

Jennifer Sinibaldi, *JPSM, University of Maryland, College Park*

Anton Örn Karlsson, *Statistics Iceland*

In the interest of better understanding total survey error, survey researchers are increasingly analyzing multiple stages of the survey lifecycle to understand how early stages might impact later ones. To contribute to this effort, this analysis examines how the repeated sampling of individuals affects nonresponse bias. Although government surveys in large countries are not likely to sample the same individual for independent general population surveys, this is a possibility when the population is small. Iceland has 325,000 people but four major surveys, including the Labor Force Survey which follows Eurostat conventions so that the data can be compared to other European countries. A review of the samples of the four surveys from 2009 to 2013 shows that over 2,000 individuals have been selected for more than one survey during this period. To determine if these individuals are more or less likely to respond, we model cooperation for the most recent survey for all sampled cases in a set time period. Upon identifying a different response likelihood for the repeatedly sampled cases, we use rich auxiliary data from the national registry to identify characteristics that may suffer from nonresponse bias. We relate these characteristics to the key variables of interest in the four major Icelandic surveys. Although Iceland is a somewhat unique case, the lessons from this analysis are broadly applicable. When all types of agencies that conduct surveys (e.g. academic, polling, market research) are considered, some individuals will be repeatedly sampled, even within large populations. The results of this analysis will allow for conservative estimates of the effect on response rates and nonresponse bias for individuals who are repeatedly sampled across different agencies.

Assessing Survey Cooperation Among Landline and Cell Phone Populations

Diane Morovati, *The Nielsen Company*

Antonia Toupet, *The Nielsen Company*

Previous research has reached a consensus that cell phone recruitment in survey research is more costly and has a higher nonresponse rate when compared to landline recruitment (Brick et al., 2007; Link et al., 2007; Qayad et al., 2013). It has been suggested that these differences in cellphone versus landline nonresponse rates depend on the type of telephone in which a person makes and receives most of their calls (Brick et al., 2006) and once contacted, does not necessarily translate into differences in the quality of responses (AAPOR, 2010). With that said, little is still known about a participant's propensity to cooperate in survey research once they have been successfully contacted by either cell phone or landline. Recent research found that upon being successfully contacted, survey cooperation (which was operationalized as survey completion) was lower among landline participants compared to cell phone participants (Qayad et al., 2013). Taken together, it seems that although it may be more challenging to reach cell phone respondents compared to their landline counterparts, they are actually more likely to cooperate once they have been successfully reached. This study further explored the notion

AAPOR 70th Annual Conference

that cell phone participants would be more cooperative in survey research compared to landline participants. Specifically, this study hypothesized that cell phone respondents would be less likely to respond “I don’t know” across a series of survey questions when compared to their landline counterparts. The data from this study came from a dual-frame probability sample of participants residing in 4,000 households. They were asked various questions regarding the type and frequency of their mobile device usage. Preliminary findings indicate that landline respondents are in fact more likely to give more “I don’t know” responses compared to cell phone respondents, $\chi^2(3, N = 8,735) = 20.57, p < .001$.

Envelope Features and Interviewer Training in a Large Advance Letter Experiment

Paul J. Lavrakas, *NORC at the University of Chicago*

Christopher Ward, *NORC at the University of Chicago*

Can Geng, *NORC at the University of Chicago*

Vince Welch, *NORC at the University of Chicago*

Benjamin Skalland, *NORC at the University of Chicago*

Jenny Jeyarajah, *Carter Consulting Inc.*

Cynthia Knighton, *Centers for Disease Control and Prevention*

There is considerable research about the effects of advance letters on survey response rates, but it does not provide definitive knowledge about how to implement them to achieve the most cost-effective outcomes. For example, very little is known about features of the envelope used in the mailing or how interviewers should utilize the letter when speaking with respondents. In Q4 2014, NORC is conducting a 3 x 3 x 2 full factorial experiment (with 18 conditions) as part of the National Immunization Survey (NIS - a very large national dual-frame RDD survey about childhood vaccinations sponsored by the Centers for Disease Control and Prevention). Three factors are tested in this experiment: (1) color of the envelope (white, tan, blue), (2) designation of the addressee (actual person’s name, Resident, Parent/Guardian) and (3) a special training curriculum to help a random subset of interviewers make more effective use of the letter with respondents. In the experiment, approximately 385,000 advance letters were mailed to addresses that could be matched to the landline RDD phone numbers in the initially designated sample. Of these addresses/numbers, approximately 21,400 were randomly assigned to each experimental condition. Hypotheses include an expectation that the non-white envelopes, the addressee indicating “Parent/Guardian,” and the special interviewer training will raise response rates. A number of dependent variables for both the screening stage of the interview and the full interview will be used in the analyses. Given the size of the experiment, the power will be in excess of 80% to detect main effect differences as small as 1%, and the power will be in excess of 80% to detect two-way and three-way interaction differences as small as 3%. Findings from this study will help address important knowledge gaps in the advance letter literature and will hopefully assist in increasing NIS response rates.

An Evaluation of the Potential to Reduce Nonresponse and Bias Using the Census Planning Database in Preparation for a Household Survey

Sarah Grady, *National Center for Education Statistics*

Cameron McPhee, *American Institutes for Research*

Celeste Stone, *American Institutes for Research*

As summarized by Groves (2006), there are characteristics of surveys (sponsorship, burden) and of respondents (gender, urbanicity) that predict response propensity. Information about response propensity from one administration of a survey can be used to develop targeted

AAPOR 70th Annual Conference

contact procedures that increase response, reduce bias, and defray costs for subsequent administrations. It is less clear, however, whether information from one cross-sectional household study can be used to predict response to a different survey. The authors evaluate whether using response propensity from the decennial Census can increase efficiency of the National Household Education Surveys (NHES) design, sponsored by the National Center for Education Statistics. Because NHES uses address-based sampling, it has been limited to address-level information, provided commercially by a sample frame vendor, for designing efficient contact strategies. In this paper, the authors evaluate the use of the Census Bureau's Planning Database (PDB), which provides block-group level response rates to the 2010 decennial census, to enhance sampling frame data and differentially target respondents. The authors conduct two analyses to evaluate the PDB for use in an NHES targeted-design strategy. First, they compare response to the NHES:2014 Feasibility Study with PDB response propensity scores by demographic characteristics such as percent minority and percent in poverty to assess whether the PDB would have accurately predicted response to the NHES:2014. Second, the authors match sampled households from the NHES:2014 study to PDB data and analyze the results of a random-assignment incentive experiment conducted in the NHES:2014 in relationship to decennial Census response propensity data. The analysis allows for a simulation of response rate outcomes, post hoc, had the PDB been used in 2014 to target likely responders with a lower incentive than the incentive given to likely nonresponders. The authors evaluate the results of the simulation on both response rates and bias in key estimates.

Experience of Multiple Approaches to Increase Response Rate in a Mixed-Mode Implementation of a Population-Based Health Survey

Urban E. Landreman, *Hennepin County Public Health Department*

David C. Johnson, *Hennepin County Public Health Department*

Mei Ding, *Hennepin County Public Health Department*

Amy K. Leite-Bennett, *Hennepin County Public Health Department*

Komal Mehrotra, *Hennepin County Public Health Department*

Emily Thompson, *Hennepin County Public Health Department*

Meghan Rosenkranz, *Hennepin County Public Health Department*

Mixed-mode data collection is of increasing interest within survey research methods, particularly when Web-based data collection has the opportunity to increase data quality and decrease cost. Address-based sampling from the U.S. Postal Service's Delivery Sequence File (DSF) presents a sustainable method to implement population-based surveys, but web-based data collection has not been extensively used in this context. Existing methods have been established to push particular modes over a series of mailings (Messer & Dillman 2011), but these methodologies often rely on monetary incentives. The project examines the performance of a mixed-mode population-based survey in a large urban setting that has the goal of providing results generalizable to the overall population. Metro SHAPE (www.MetroSHAPE.us) is an ongoing population-based health surveillance project administered collaboratively by eight health departments in the Minneapolis-Saint Paul, Minnesota area. In 2014, a random sample of 58,000 households were selected to participate in the survey using the DSF as the sampling frame. It is anticipated that 10,000 individuals will complete the survey. Selected households were twice offered a chance to complete the survey on-line using an assigned ID and password. They subsequently were sent a paper version of the survey. Some participants were given a small non-monetary incentive with the initial mailing and others were not. The presentation will discuss: 1) the impact of strategies used to increase the inclusion and participation of persons in low-income households, racial and ethnic minority groups, and households without computers; 2) the differences in demographic characteristics of survey respondents between the on-line

AAPOR 70th Annual Conference

and paper modes and potential mode effects; 3) the impact of non-monetary incentives; and 4) the benefits and challenges of administering a large population-based health surveillance project jointly with multiple government agencies. The survey went into the field in September 2014 and is scheduled to finish in early 2015.

Friday, May 15
10:00 a.m. – 11:30 a.m.
Concurrent Session D

Mini-Conference: Innovation in Federal Surveys - Opportunities, Progress and Challenges

Mini-Conference: Innovation in Federal Surveys - Opportunities, Progress and Challenges

Peter V. Miller, *U.S. Census Bureau*

Federal surveys, which provide important benchmarks for public opinion and survey research practitioners, are subject to the same economic and societal forces that affect state government, commercial and academic research enterprises. Flat or declining budgets, falling rates of survey cooperation and demands for more and faster release of data pose significant problems for federal survey managers. Innovation in survey practice is essential to address these issues. This panel is comprised of five papers that present research on change in a range of federal surveys. The first paper provides an overview of opportunities for, and challenges to invention in governmental survey research. It offers a context for the four succeeding papers. Three papers to follow provide a look at opportunities and progress in data collection innovation. Two papers present research on adaptive survey design, at different stages of maturity, in different education surveys - one a household survey (the National Survey of College Graduates) and one a study of schools (the National Principal and Teacher Survey). A fourth paper discusses the use of alternative indicators of data quality in tandem (R indicators and response rate) to track data collection progress in the Annual Survey of Manufacturers. The final paper discusses challenges to innovation posed by interviewer noncompliance with new case management methods in face-to-face household surveys - research for the 2020 Decennial Census and research in the National Survey of Family Growth. Taken together, the papers provide a useful perspective on efforts in federal surveys to respond to the economic and societal challenges facing all survey researchers.

Innovating in Federal Surveys: Invention's Mother Meets "The Way We Do Thing Here"

Peter V. Miller, *U.S. Census Bureau*

Despite the broad recognition among Federal survey managers that established data collection procedures cannot be sustained due to rising costs, falling survey budgets and declining respondent cooperation, adopting new methods is not straightforward. The weight of precedent, the aversion to disrupting time series and the enormous importance of the statistics produced in these surveys provide ample reason to try to preserve the methods that undergird them. New

AAPOR 70th Annual Conference

procedures intended to alleviate cost and data quality pressures require careful vetting. New perspectives on survey quality, which may offer more useful goals for survey management, must be proven and related to traditional measures. Cost measurement must be improved and tailored for daily survey management. Systems that can support new management methods must replace the current IT infrastructure. Major change in incentives for field staff is required so that new case management methods can be tested. These essential steps and others must be made in an environment in which claims on resources for innovation compete against claims for resources just to keep surveys going. Arrayed against the conditions favoring the status quo are strong change leadership, opportunities for innovation in survey management opened up by access to administrative records data, paradata and accumulated response data, a culture that supports risk-taking and technical expertise to design and implement research to test new ideas. Despite the considerable power of inertia, these factors offer promise that "Invention's Mother will win out over 'The Way We Do Things Here.'"

2015 National Survey of College Graduates: Enhancing the Use of Adaptive Design

John Finamore, *National Science Foundation*

Stephanie Coffey, *U.S. Census Bureau*

Benjamin Reist, *U.S. Census Bureau*

A major goal of adaptive design is to increase the efficiency of survey operations, with respect to cost and timeliness, while maintaining or improving survey data quality. In 2013, the National Survey of College Graduates (NSCG) included an adaptive design methodology study to explore whether data monitoring techniques and data collection interventions could be implemented during the data collection effort of a large-scale national survey. The 2013 NSCG adaptive design study showed that it is possible to monitor data as it is being collected and use the monitoring results to inform data collection interventions that have potential to improve data quality and reduce survey costs. Now that our ability incorporate adaptive design techniques into a large-scale national survey has been established through the 2013 NSCG adaptive design study, our current goal is to enhance our use of adaptive design. With that goal in mind, the 2015 NSCG will include an expanded adaptive design methodology study that includes a larger sample and, taking advantage of the longitudinal nature of the NSCG data, explores the use of adaptive design techniques on both new and returning sample members. The 2015 NSCG study will explore monitoring metrics and data collection interventions in an effort to address the survey issues most important for the NSCG including nonresponse bias, data timeliness, and survey costs. The 2015 NSCG adaptive design study will employ a sample size that provides the statistical power necessary to make more definitive statements about statistical differences between the study group and the control group on various measures, including response rates, R-indicators, cost, and effect on key estimates. This presentation will discuss the development of the 2015 NSCG adaptive design methodology study, the consideration and determination of data monitoring metrics, data collection interventions and the initial study results.

AAPOR 70th Annual Conference

Adaptive Design for the National Teacher Principal Survey

Minsun Riddles, *Westat*

David A. Marker, *Westat*

Louis Rizzo, *Westat*

Erin Wiley, *Westat*

Andrew Zukerberg, *National Center for Education Statistics*

Statistical agencies are frequently confronted with the trade-offs between timeliness (relevance) and accuracy. Waiting for the last responses and quality reviews can improve accuracy but delay production of the data sets and analyses, reducing their relevance to users. The National Center for Education Statistics has conducted the quadrennial Schools and Staffing Survey (SASS) since the 1980s. Beginning with the 2015-16 school year SASS will be replaced with a new biennial National Teacher Principal Survey (NTPS). As part of the design for the NTPS, we are reviewing response patterns and paradata collected during the 2011-12 SASS to develop an adaptive design for the new study. Adaptations may include when to switch data collection modes, when to stop overall data collection, and revisions to methods for contacting respondents. We will therefore simultaneously examine multiple components of Total Survey Error, including nonresponse bias, mode effects, and relevance (time lag from reference period to publication). This presentation will discuss the different approaches considered in the analysis and provide a framework for other studies considering adaptive design approaches.

Responsiveness and Representativeness in an Establishment Survey of Manufacturers

Eric Fink, *U.S. Census Bureau*

Joanna Fane Lineback, *U.S. Census Bureau*

Response rates, because of their ease of calculation and understanding, traditionally have been used as data-collection-quality metrics. However, recent research has cautioned against solely relying on response rates, as survey programs' efforts to increase these rates may lead to increasing the likelihood of biasing survey estimates. R-indicators have been proposed as a complementary measure that can give insight into the data collection process that response rates alone cannot explain (Schouten and Cobben 2007). In this paper, we calculate traditional response rates and R-indicators for the 2011 Annual Survey of Manufacturers and demonstrate that, when used in tandem, they can give a more complete picture of the data collection process, particularly the nonresponse follow-up. In particular, we show that despite increasing response rates during nonresponse follow-up, representativeness across important design variables, such as establishment size, decreases, owed, in part, we hypothesize, to concentrating follow-up on those establishments expected to contribute the most to total estimates. This lack of representativeness is a possible source of bias in resulting survey estimates if nonresponse adjustments do not correct for over or underrepresented areas. Hence, we also examine the R-indicator post-nonresponse imputation. We discuss the tradeoff of reducing sampling variability versus reducing nonresponse bias. Further, we incorporate associated costs into our analysis, and discuss how these cost/quality indicators can be used in conjunction with data quality metrics to provide a more complete picture of the efficacy of the data collection process.

Challenges to Innovation in Face-to-Face Surveys Posed by Interviewer Noncompliance

Gina K. Walejko, *U.S. Census Bureau*
James Wagner, *University of Michigan*

Traditionally, interviewers working in computer-assisted personal interviewing (CAPI) surveys have a great deal of flexibility in organizing their work, including allowing them to choose when to work, which cases to contact, and what time of day to make calls. Recent studies testing survey innovations have requested CAPI interviewers to follow explicit instructions intended to increase data quality and promote efficiency. In some instances, interviewers follow procedures, while, in others, they do not comply with such protocols. We discuss the challenges associated with implementing innovative field procedures reliant on CAPI interviewers and report on interviewer compliance with a variety of requested actions. We use contact history paradata and interviewer debriefings data from two decennial census field tests and the National Survey of Family Growth to examine CAPI interviewer compliance with several procedures, including filling out contact history information accurately, working specific “prioritized” cases, working cases at recommended times, seeking proxy interviews for specific cases, and following proposed work schedules. We discuss the circumstances under which interviewers complied and did not comply with requests for specific actions. We show that interviewers attempt a compromise between what is convenient for them and what they have been instructed to do. We conclude with suggestions to survey practitioners who are designing data collections that rely on CAPI interviewers’ to comply with innovative protocols.

The Economy, Engagement and Political Participation

Grievance Asymmetry in Economic Voting and Voter Sophistication

Ju Yeon Park, *New York University*

Previous literature suggests that voters tend to punish an incumbent when the economy is bad more than they reward him for a good economy, and this bias is called grievance asymmetry. In this project, I examine whether this behavioral pattern in economic voting is conditional on an individual voter’s level of political knowledge. I borrow three different theories explaining individuals’ stronger reactions to negative stimuli from the fields of political communications, economic behavior, and psychology and develop them into three alternative hypotheses on voting behavior of sophisticated voters and naive voters in response to economic conditions. I use the Comparative Study of Electoral Systems after merging multiple modules covering 43 countries and 103 elections with 62,409 individual responses. In order to overcome the problem of endogeneity between survey respondents’ economic evaluations and their support for incumbents in survey data, I use hierarchical models that substitutes survey responses on economic evaluations with objective economic indicators. I find that sophisticated voters tend to be tougher when evaluating their incumbents at times of both economic booms and busts, which gives supports for the hypothesis that voters’ consumption of news that tend to report negative news more frequently than positive news may account for the pattern of grievance asymmetry in economic voting.

AAPOR 70th Annual Conference

Improving on the Standard Prospective Economic Evaluation Question

Marco Morales, *New York University*

Economic voting is perhaps one of the most studied relationships in Political Science. The last 30 years of research on economic voting based on survey data have employed a question that is ill-suited to study the relationship between the economy and vote choice in a manner that would be consistent with the theory. Focusing on prospective sociotropic economic assessments (i.e. evaluations about the future state of the national economy), the paper dissects the question that is typically used in survey-based economic voting analyses and singles out two specific causes of measurement error: the phrasing and the response scale of the commonly used question. It also proposes a corrected survey instrument - based on survey experiments conducted in the U.S. and Mexico - that consists of a battery of questions that condition on each candidate being in office, and utilizes response scales that disambiguate what respondents mean when they respond that the economy is the same as it is now.

Public Support for a Balanced Federal Budget

Andrew W. Crosby, *University of Illinois at Chicago*

Allyson L. Holbrook, *University of Illinois at Chicago*

Requiring a balanced budget at the federal level in the U.S. would mean that the federal government could not spend more than its revenues and available resources. A number of countries have balanced budget requirements at the federal level and 49 states in the U.S. also have some form of law that requires that the state budget be balanced. A federal-level balanced budget requirement, up to and including an amendment to the U.S. Constitution, has been considered in the U.S. as early as 1798. Using data from more than 40 national public opinion polls, this paper examines trends in public opinion toward the requirement for a balanced federal budget in recent decades, including support for a balanced budget amendment to the U.S. Constitution. In addition, we develop a predictive model of support for a balanced budget amendment based on a variety of factors including respondent level variables (e.g., education and party identification), contextual factors (e.g., financial conditions, president) and question wording (e.g., the extent to which support for balancing the federal budget is reduced the question points out possible consequences such as reductions in spending for programs such as Social Security).

The Nitty Gritty: The Relationship Between Non-Cognitive Skills and Civic Engagement

Steven Andrew Snell, *Duke Initiative on Survey Methodology, Duke University*

D. Sunshine Hillygus, *Duke University*

John Holbein, *Sanford School of Public Policy, Duke University*

The robust empirical relationship between education and political participation has long sparked debate about the extent to which intelligence and cognitive skills determine engagement in civic and political life. In this paper, we consider the previously overlooked role of non-cognitive skills, especially grit or perseverance, as a predictor of political participation. A growing body of research has shown that non-cognitive skills may be as important as cognitive skills in predicting success in school and the labor force. This project builds on this literature by examining if grit is a meaningful determinant of engagement in civic and political life. We argue that grit shapes the extent to which individuals are willing to overcome the costs of political participation. Drawing on data from the Adolescent Health Longitudinal Survey, a large national

AAPOR 70th Annual Conference

panel survey of youth as they transition into adulthood, the 2014 Cooperative Congressional Election Study, and an MTurk question wording experiment, we examine the measurement properties of grit and evaluate the relationship between grit and political participation, above and beyond traditional cognitive skills measures. As part of this project, we evaluate the reliability and validity of measures of grit, including stability of grit measures over time. Initial analyses suggest that grit is not only a strong predictor of political participation, but it can be shaped by educational policies aimed at adolescents. These findings have important implications for theories and models of political participation and suggest new policy avenues for increasing the engagement of young citizens.

Prospective Economic Evaluations: Do They Measure What We Think They Do?

Marco Morales, *New York University*

Measures of perceptions of the future state of the economy have fed a plethora of studies on the impact of economic conditions on vote choice. Yet, the phrasing of the standard economic prospective evaluation question makes it difficult to determine what information are individuals providing when they answer it. Specifically, when respondents are asked about the future state of the economy in pre-election surveys (while the winner of the election is unknown), it is unclear whether respondents provide an assessment of the future state of the economy under i) their preferred party; ii) under the likely winner; iii) under the current incumbent; or iv) under some other alternative. This paper seeks to explain what is contained in the answers to the standard prospective economic evaluation question. Evidence from the 2005 British Election Studies (BES) suggests that individuals answer this question with an assessment about the future state of the economy under their preferred party with some qualifications: the incumbent party if they prefer that party (regardless of whether they think it will win the election), or a challenger when they prefer that party and think it is also the likely winner of the election. No clear results are found for individuals who prefer a party that they think is unlikely to win. After the election, respondents converge answering about the future state of the economy under the party that won the election. These findings stress the need to replace the current question with a battery of questions that capture the future state of the economy if each one of the parties competing were to win the election. This information would provide the necessary information to test economic voting more adequately.

ABS Refinements

Exploring the Feasibility of Conducting a Two-Stage Mail Survey in a Single Stage

Rebecca Medway, *American Institutes for Research*

Danielle Battle, *American Institutes for Research*

Address-based mail surveys have become increasingly popular in recent years. A challenge in conducting such surveys in the U.S. is determining which households are eligible for the survey, and then ensuring that the appropriate household member(s) complete it – all without the aid of the interviewer that would be present in other modes to help facilitate this process. A popular option for addressing this challenge is to first send a screener questionnaire to identify eligible households and the eligible individuals living in them; eligible households are later sent the full questionnaire and told which household member(s) should complete it. Drawbacks to such two-stage designs include the time and cost associated with sending the two stages and the potential for lower response rates due to challenges in getting households to respond to both stages. We report on efforts to combine the household and individual-level eligibility

AAPOR 70th Annual Conference

determination and topical survey data collection stages into a single-stage. As part of a national pilot study of adults conducted in 2013, sampled households were mailed questionnaires that included a “screener” page in which the sampled households were asked to indicate the number of eligible individuals in the household. Instructions were included requesting that each of these eligible individuals complete one of the three survey questionnaires that had been mailed to the household. Households were randomly assigned to receive the questionnaires as one composite booklet or three individual booklets. This presentation will compare results from this single-stage design overall to the two-stage design typically employed by this survey on outcomes such as the response rate, screener responses, respondent demographics as compared to frame data and the Current Population Survey, responses to key questions, and response quality. We will also report on differences between the composite- and individual-booklet conditions in terms of these same outcomes.

Tracking and Evaluating Updates to the ABS Frame Over Time

Jennifer Unangst, *RTI International*

Joe McMichael, *RTI International*

Address-Based Sampling (ABS) frames are usually based on the USPS Computerized Delivery Sequence (CDS) file, which is updated by the USPS twice monthly. With each update, addresses are added, deleted, revised and/or moved between the CDS and No-Stat files. It is important for researchers to understand the frequency and scope of these changes, so they can assess the quality and coverage of their frame. This paper tracks more than 130 million addresses from the national CDS file over time from 2010 to present. We detail where and when changes were made, as well as the types of changes (e.g., whether a change was an addition, deletion, or vacancy status change), across sub-areas of interest for which ABS coverage rates differ (e.g., rural areas). We additionally explore the possibility of using this information to identify new construction and to update census-based housing unit counts which are often used as sampling size measures. This analysis has never been available to the public and will aid researchers in study planning where ABS frames are being considered.

Elements of Non-Response Across Mode in ABS Surveys: The Use of GIS-Based Modeling to Understand the Importance of Space

Ned English, *NORC at the University of Chicago*

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*

Multi-mode address-based (ABS) studies have emerged in recent years as a potential solution to declining response and coverage in telephone frames, driven both by improvements in technology and the availability of comprehensive residential address frames. In such studies respondents are usually given the option of responding by telephone, web, mail, or in-person with order depending on design. This study investigates how geography influences survey response rates by mode, with special attention to socio-economic population groupings and the urban-rural continuum. Because ABS studies are fundamentally geographic, it is possible to use capabilities of modern geographic information systems (GIS) to characterize spatial impacts of mode. For example, we may expect geographic clustering of web responses due to internet penetration; such clustering may introduce additional sources of error in analytical results. We use data from a multi-mode health survey in Illinois to extend previous research on web-only response rates by quantifying spatial elements of web, telephone, and mail responses, as well

AAPOR 70th Annual Conference

as telephone-matching in a three-way comparison. We show spatial differences in response-rates and substantive questionnaire data using Self-Organizing Maps, Spatial-Lag modeling, Moran's I, and local indicators of spatial autocorrelation (LISA). In so doing we describe the geographic and demographic characteristics of places that tend to favor a given mode, how response rates cluster geographically, and how mode, in combination with geography and demography, correlates with substantive survey results. Our results are useful to practitioners of multi-mode ABS surveys who want a more detailed view of the impact of particular designs on health surveys.

Can I Get Your Phone Number? Examining the Relationship Between Household, Geographic and Census-Related Variables and Phone Append Propensity for ABS Samples

Kristen Olson, *University of Nebraska-Lincoln*
Trent D. Buskirk, *Marketing Systems Group*

Address-based samples afford survey researchers with auxiliary data that can be linked directly to an address via its geographic coordinates. These variables can assist in designing samples and nonresponse adjustments, or, increasingly, in adaptive survey designs. Examples of appended variables include household level variables like age of householder, presence of children or Hispanic surname and aggregate income of the block group or percentage of the block group that are Hispanic. One common design that uses this information is the ABS Sample with Phone Follow Up option in which a sample is selected from an ABS sampling frame and phone numbers are appended to as many sampled addresses as possible. In the field, mailings are sent to all sampled addresses, and follow up attempts include phone calls to addresses with phone numbers and mailings to addresses without phone numbers. Survey completion rates and characteristics of respondents have been reported to differ for addresses with a phone match compared to addresses without a phone match. Yet more work is needed to understand if those addresses with phone match are a systematically different subset of the overall sampling frame, rather than simply differences in characteristics of people who respond versus do not respond to each mode. In this study we investigate the relationship between over 500 household, geographic and census related variables and the likelihood of an address to have a phone number matched to it using an ABS sample of 1,000,000 records randomly selected from a U.S. residential ABS Sampling frame. Overall, about 45% of households have an appended phone number. We will present detailed profiles of those addresses for which a phone number can and cannot be linked. Implications of this research for tailored designs using mixed ABS and phone and nonresponse adjustments will also be discussed.

DPV Codes and Response Rates in the National Children's Study

Rachel Carpenter, *NORC*
Erin Tanenbaum, *NORC*
Lauren Bishop, *NORC*

Research has shown diminishing response rates in mail, telephone, and in-person surveys are greatly threatening the reliability and generalizability of survey results. Maintaining high participation and response rates is particularly important for longitudinal studies such as the National Children's Study (NCS) in which attrition tends to increase over time. For the NCS, mail is an important way to communicate with and track participants, to complete mail surveys and in-person specimen collections, and to reduce respondent burden by linking existent data related to neighborhood characteristics. For these reasons an accurate mailing addresses is crucial for all participants. In this paper we explore the use of Delivery Point Validation (DPV)

AAPOR 70th Annual Conference

codes and residential address flags as a prioritization method to investigate and hopefully correct potentially poor addresses. Between 2002 and 2007, the United States Postal Service introduced Delivery Point Validation (DPV) codes and Residential Delivery Indicator (RDI) to identify potentially inappropriate addresses prior to mailings. DPV codes help validate accurate delivery address information by identifying inaccurate, incomplete, or erroneous addresses. In addition to DPV codes, an RDI code is readily available to mailers to indicate whether an address is residential, commercial, or rural. These tools are used to comply with USPS pricing guidelines for mass mailers but may also be useful for survey methodologists, communications teams, and statisticians to validate addresses. This paper provides an analysis of whether DPV codes and RDI increase address quality in a longitudinal study. In addition, this paper builds on previous research that suggested that mailings to low-income respondents are more likely to be returned to the survey organization as undeliverable. This analysis will examine whether DPV codes and RDI may be helpful in reducing the discrepancy between undeliverable rates for low-income versus high-income respondents.

Innovations in Reducing Respondent Burden and Fatigue

A Data-Driven Evaluation of the Burden and Benefits of the Questions Included in the American Community Survey

James B. Treat, *U.S. Census Bureau*

Gary B. Chappell, *U.S. Census Bureau*

Sarah K. Heimel, *U.S. Census Bureau*

Todd R. Hughes, *U.S. Census Bureau*

Given the mandatory nature of the American Community Survey (ACS) it is especially important to ensure that the survey's questions do not overburden the public, but what factors should determine which topics and questions should be included? Periodic reviews are necessary to ensure that there is clear and specific authority and justification for each question on the ACS, the ACS is the appropriate vehicle for collecting the information, the survey minimizes the burden placed on its respondents, and the quality of the data from the ACS is appropriate for its intended use. In 2014 the Census Bureau examined all 72 questions contained on the 2014 ACS questionnaire and conducted its most comprehensive effort with the highest participation ever undertaken to establish the legal authority and justification of every question on the ACS. This session examines in detail the methodology followed to analyze each of the 72 questions on the 2014 questionnaire. We begin with a description of the data inputs associated with the 19 pre-specified criteria used for the content review – 13 related to benefits, and six related to costs or burden. Then we discuss how the benefit and cost criteria were weighted and applied based on the recommendations of the federal statistical body chartered to provide such guidance to the Census Bureau, the Interagency Council on Statistical Policy Subcommittee for the ACS. We then present the results of taking each ACS question through this analysis, including a detailed examination of those items recommended as potential candidates for removal. Finally, we present current and proposed lines of research to examine other sources that could provide alternatives to information currently collected by the ACS, and how existing questions might be modified to reduce survey burden further in the coming years.

Respondent Burden: Lessons Learned from a Mixed-Methods Approach to Assessing Measures of Perceived Burden

Morgan Earp, *Bureau of Labor Statistics*

Scott Fricker, *Bureau of Labor Statistics*

The U.S. Consumer Expenditure Quarterly Interview Survey (CEQ) collects data on respondents' perceptions of burden using a set of nine items administered at the end of the final wave of data collection. Previous research using these items has focused on relationships between respondents' reported burden, their characteristics, and reporting behaviors, and compared these subjective burden measures with more traditional, "objective" burden indicators (e.g., interview length). In this study, we use Confirmatory Factor Analysis (CFA) and Item Response Theory (IRT) to evaluate the psychometric properties of the existing CEQ burden items. We also report the results of lab- and web-based qualitative examinations of respondents' interpretations of these items versus alternative versions. We compare findings from these analyses highlighting the benefits of mixed-methods research, discuss the impact of item response scales, and suggest options for improving assessments of respondent burden.

Current and Future Life Evaluation: Results of Experiments in the Gallup World Poll to Reduce the Length of the Question Wording

Robert D. Tortora, *Gallup*

Sofia P. Kluch, *Gallup*

Ken Kluch, *Gallup*

This paper presents the results of a large-scale experiment, conducted using the Gallup World Poll, to reduce the length of the question wording when asking about life evaluation. Currently Gallup uses the Cantril Self-Anchoring Striving Scale (Cantril, 1965), often called the ladder of life, to ask about current and future life evaluation. Respondents are first asked about current life evaluation and then asked about expected life evaluation, that is, life in 5 years. In this experiment the wording of the question stem is reduced from 147 words to 87 words, cutting the time an interviewer takes to read the question in about one-half. The response scale on both the Cantril ladder of life and the proposed wording is an 11 point scale, zero to ten. The new wording (Tortora, 2014) was tested in nationally representative face to face surveys conducted in Congo Brazzaville and Ethiopia in 2012. No differences were found in these two experiments for both current life evaluation and life evaluation in five years. In 2014 Gallup extended the testing to over 60 countries around the world, including countries in Africa, Asia, Europe, and the Americas. Both RDD telephone (including mobile phones) and face to face surveys are included in the experiment which randomly assigned participants to one of the two forms of the question wording. Almost all of the surveys in the experiment are based on nationally representative samples. About 1/3 of the surveys are conducted by telephone and the remainder by personal interview. Besides an overall comparison of the results, the data are compared by the gender, age and education level to examine effect of the change in question wording.

Can Survey Instructions Relieve Respondent Burden?

Erica C. Yu, *Bureau of Labor Statistics*

Survey designers must make choices when developing surveys. Many of these choices are made to reduce respondent burden. For example, survey designers sometimes use screener questions to minimize the number of questions that a respondent is asked. However,

respondents are generally not aware of these design choices and can only judge their survey experience based on the questions they received, without the knowledge that the survey could have been longer. If survey designers could control the reference set by which respondents judged their burden, they may be able to improve the experience of participating in the survey. Research has shown that judgments are often made relative to a reference set, and that different reference sets can lead to different judgments. The reference set may be determined at the time of judgment, such as by asking the respondent to rate their experience compared to preparing taxes (less burdensome) or ordering pizza (more burdensome). Or, the reference set may be determined during the experience, such as by a notification that the survey will be longer (less burdensome) or shorter (more). This study tests whether managing the reference set during the survey experience using instructions affects burden ratings and data quality. Participants were recruited from Amazon Mechanical Turk to complete an online survey (n=360). Two factors were manipulated in a between-groups design: actual number of questions asked (16, 32) and screening outcome (screened in, screened out, no screener instructions). The participants who received no screening instructions serve as a control group for understanding the perceived burden of the surveys without the impact of the experimental instructions. After completing the survey, participants answered questions to inform multiple measures of burden that have been proposed in the literature. Discussion will include whether survey instructions can alter perceptions of burden and other survey outcomes.

The Effects of Total Navigational Burden, Length of Instrument and Page Complexity on Item Non-Response

Josephine P. Leonard, *Experian Marketing Services*

Max F. Kilger, *Experian Marketing Services*

Jennifer M. Bouterse, *Experian Marketing Services*

Respondent fatigue has been a concern for survey researchers for many years. In particular, respondent fatigue for long surveys (> 1 hour in length) can have a debilitating effect on the ability of researchers to collect a large number of measures from a single source at a single point in time. There are a number of factors that contribute to respondent fatigue including length of survey, the complexity of the measures and topics involved, the physical layout of the survey, the visual complexity of the survey page, the number of required responses per page and the magnitude of micro-tasks such as navigation that must take place on each page of the survey. While respondent fatigue is often addressed anecdotally in research articles and there are some studies that examine respondent fatigue in the form of repeated administration of instruments, there is a dearth of studies that examine, especially on an empirical level, some of the factors that contribute to respondent fatigue and particularly within large survey instruments. In this study we utilize a very large survey instrument – a 130 page survey booklet – that is used in a large sample (n~25,000) national probability study. In particular, we examine item non-response rates for a large number of gate questions (n~180 questions) where a yes or no answer is expected from the respondent. A multivariate model predicting whether or not a respondent responds to a gate question on a page will be constructed and will utilize as independent variables an estimate of the total navigational burden for the page, the mean number of responses across respondents for the page as the indicator of response burden, the page number as an indicator of the effect of length of instrument and a measure of the visual complexity of the page.

Race: Attitudes, Measurement and Effects

Understanding the Roots of Attitudes on Inequality

Damla Ergun, *Langer Research Associates*

Julie Phelan, *Langer Research Associates*

Gary Langer, *Langer Research Associates*

What drives people's attitudes on inequality and related issues, including which groups are deserving of help, who among the public is motivated and willing to take action to provide that help, and what policies should be devoted to addressing inequality? These questions guided a comprehensive research effort commissioned by Opportunity Agenda. Regression analyses identify the strongest predictors of seeing unequal treatment of social, racial and ethnic groups as a serious problem, and willingness to take action to address inequality. Results point to the importance of personal experiences, ideology and beliefs about sources of inequality in society, among others, in shaping these attitudes. As a next step in understanding motivations to act on social issues, cluster analyses identified segments of the U.S. population on the basis of their views on discrimination, their personal experiences with it and their willingness to take steps to address it. Six typologies emerge, with differing demographic characteristics, policy views and core values that summarize the constellations of attitudes and behaviors on equality issues and related social policies

Developing Innovative Methods for Community-Based Research: The Toronto Black Experience Project

Keith Neuman, *The Environics Institute for Survey Research*

Much of the survey research world today is focused on adapting to rapidly changing online technology and social media, and to the evolving norms around communications and privacy. But there continues to be an important role for traditional modes of survey research, such as in-depth interviewing that takes place face-to-face where people live. In some cases this type of methodology is the only viable choice when the requirement is for in-depth and often sensitive information from specialized populations. The Black Experience Project is a groundbreaking community-based survey research project with the Black community in the Greater Toronto Area, with the goal of better understanding the opportunities and challenges of this population, and contributing to effective policies and improved outcomes. The Black population numbers over 400,000 and encompasses a highly diverse set of communities, with individuals coming from more than 100 countries, and including both recent immigrants and sixth generation Canadians whose predecessors came north as part of the underground railway. The study is unique in its focus on the "lived experience" of this population (exploring identity, values, aspirations and experiences), rather than the prevailing research focus on specific problems, such as housing, health and poverty. A unique community-based research model was developed to address the various challenges of this project, among which include: a) sampling to include the full range of this diverse population; b) gaining the support from a community skeptical (if not suspicious) of research that to date has done little to benefit it; c) actively involving individuals from the community in all phases of the research; and d) building broader community support to ensure the study findings lead to meaningful improvements and empowerment. The paper will provide an overview of the research approach and methodology, and present selected findings that illustrate the power of the methodological approach.

The Booker Effect – The Sequel: A Comprehensive Investigation of Race of Interviewer Effects in New Jersey U.S. Senate Elections

Ashley A. Koning, *Rutgers University*

David P. Redlawsk, *Rutgers University*

Cory Booker (D) – the newly elected African-American junior U.S. senator from New Jersey and former Newark, NJ Mayor known for his tweeting and headline-making – faced his first re-election campaign in Fall 2014. Much like the previous year, Booker was expected to easily beat Republican rival Jeffrey Bell – who last made a major run in New Jersey politics as the Republican U.S. Senate nominee in 1978. Booker held a double-digit lead in statewide polling leading up to Election Day, when he won by just over 13 points. While not publicly polling on the Senate in 2014, the Rutgers-Eagleton Poll embedded several experiments to understand what might impact voter responses to Booker as an African-American candidate. Earlier analysis of the Poll's significant 2013 "miss" suggested it suffered from a "race of interviewer" effect due to the diversity of its interviewers, drawn from the majority-minority Rutgers University student population, in addition to other issues. The 2014 race provided the opportunity to further investigate this phenomenon through several survey experiments and extensive data collection on name recognition of the candidates, feeling thermometers, candidate affect, racial resentment batteries, and race and ethnicity of interviewers (both respondent perceptions and actual race and ethnicity of interviewer). We report the results of this investigation, which appears to be the most comprehensive attempt to examine race of interviewer/race of candidate effects to date. In addition, we compare the 2014 general election effects with the 2013 special election. Results suggest the continued potential for complex race of interviewer effects, having important implications for African-American candidates, live interviewing, social desirability bias, and telephone interviewer populations.

Examining Acquiescent and Extreme Response Styles between Face-to-Face and Web Surveys

Mingnan Liu, *SurveyMonkey*

Frederick Conrad, *University of Michigan*

Sunghee Lee, *University of Michigan*

Likert scales are popular for measuring attitudes, but response style, a source of measurement error associated with this type of question, can result in measurement bias for this particular type of question. This study investigates the effect of data collection mode on both types of response styles using the 2012 American National Election Studies (ANES) data. 2012 was the first year in which ANES conducted two parallel surveys, one through face-to-face interviews and another through Web, using two independent nationally representative samples and an identical questionnaire. We used three sets of balanced Likert scales from the survey to measure the acquiescent and extreme response styles. Using the latent class analysis approach, we find that: 1) both acquiescent and extreme response styles exist in both face-to-face and Web survey; 2) face-to-face respondents demonstrate more acquiescent and extreme response styles than Web respondents; 3) the mode effect on extreme response style is larger for black respondents than for white and Hispanic respondents. For acquiescent response style, however, the mode effect is similar across the three racial/ethnic groups.

Uses of Paradata

Effective Observational Strategies for Face-to-Face Survey Interviewers

Brady T. West, *Institute for Social Research, University of Michigan-Ann Arbor*

Dan Li, *The Search Agency*

Yimeng Ma, *Michigan Program in Survey Methodology, University of Michigan-Ann Arbor*

In face-to-face survey data collections, observations that interviewers record describing selected features of all sampled units represent a promising, cost-effective source of paradata that survey organizations can use for nonresponse adjustment of survey estimates and responsive survey designs. Recent research indicates that interviewers can successfully observe and record features of sampled units that are correlated with key survey measures and response propensity. However, these observations can be error-prone, and additional research has demonstrated that errors in these observations can limit the effectiveness of nonresponse adjustments. Furthermore, initial studies have found that interviewers vary substantially in the accuracy of their observations, even after controlling for interviewer- and household-level covariates. No existing study has attempted to identify additional sources of this variance. Motivated by psychological and anthropological literature examining sources of bias in human observation and a pilot study of observational strategies presented at the AAPOR annual conference, this study performed in-depth qualitative analyses of more than 50,000 justifications provided by interviewers in the National Survey of Family Growth (NSFG) for their observations on two key features of all sampled NSFG households that can be easily verified: the presence of children under the age of 15, and expected probability of household response. Cluster analyses of interviewer-level tendencies evident in the justifications suggest that unique subgroups of NSFG interviewers do exist based on the observational strategies used, and these strategies are not necessarily dictated by features of the areas in which the interviewers are working. Multilevel models comparing the accuracy of the observations among the identified subgroups of interviewers while controlling for other relevant correlates of observation accuracy suggest that alternative strategies employed by the interviewers are more or less effective for improving the quality of the observations, and these results have direct implications for standardized training of interviewers on these tasks.

Investigating Respondent Multitasking in Web Surveys with Paradata

Anže Sendelbah, *University of Ljubljana*

Multitasking refers to concurrent or sequential combinations of activities. While cognitive sciences have been elaborating on multitasking since the 1930s and media multitasking (multitasking involving media activities) has been extensively researched across different disciplines in recent decades, this behavior is rarely discussed in the context of survey methodology. Respondent multitasking (RM) happens when a respondent does any secondary activity while responding to a questionnaire. RM has been theoretically associated with worse response quality (e.g. Holbrook, Green, & Krosnick, 2003; Lynn & Kaminska, 2013) and scarce empirical evidences from telephone surveys have partly supported this assumption (Kennedy, 2010; Lavrakas et al., 2010). Studies on telephone and web surveys (Zwarun & Hall, 2014) predominantly use self-reports to measure RM. In such studies, reported shares of respondents who had engaged in RM range between 30 (for web surveys) and 50% (for telephone surveys). However, people tend to underestimate the amount of time spent on multitasking (e.g. Iqbal & Horvitz, 2007). While using paradata to measure RM (e.g. Stieger & Reips, 2010) could circumvent this issue, more investigation is needed to refine this approach. We bring new

AAPOR 70th Annual Conference

insights to under-researched topic of RM in web surveys and present a new way of measuring it via advanced paradata (Vehovar et al., 2012). We have developed two different sets of RM indicators: (a) focus-out events (indicating whether a respondent has switched to another window or tab); and (b) time-out events (indicating whether a respondent has spent a substantially longer time to complete a questionnaire subpage). Preliminary results from a case study (n=267) show that 56% of respondents had done some form of RM. While focus-out indicators and time-out indicators are significantly correlated, there are also some considerable differences. Further, we used a negative binomial regression and found a relationship between RM, questionnaire complexity, and item nonresponse.

Determining Potential for Breakoff in Time Diary Survey Using Paradata

Douglas Wettlaufer, *University of Nebraska-Lincoln*

Hariharan Arunachalam, *University of Nebraska-Lincoln*

Gregory Atkin, *University of Nebraska-Lincoln*

Adam Eck, *University of Nebraska-Lincoln*

Leen-Kiat Soh, *University of Nebraska-Lincoln*

Robert F. Belli, *University of Nebraska-Lincoln*

Knowledge of whether a respondent is going to breakoff from a survey is usually determined when the event happens which is too late for adjustments to keep the respondent engaged. During the course of a survey, respondents perform actions resulting in either a successful or non-successful survey; thus, a survey can be thought of as a sequence of actions that are being recorded as paradata. From the paradata sequence, useful derived statistics could be response speed, how often they change answers, or number of responses by a certain point in the survey. By analyzing the paradata, we aim to predict whether a respondent will breakoff. In this work, paradata collected by the instrument used in the American Time Use Survey (ATUS), administered by the Census Bureau, is used to make predictions on the outcome of a survey. The data set used is from 2010 and contains 347,722 successful cases and 509 breakoffs. Sequential learning is applied to this data to use attributes of the survey for predicting outcomes such as if the respondent will breakoff. For instance, if the time to enter response data begins to slow down that may indicate the respondent is losing interest and may quit prematurely. Another possible application of this method would be to predict whether the respondent would be an outlier due to their current actions. In this context sequential learning is a pattern recognition task that will classify each sequence of actions within a survey as either leading to a breakoff or not leading to a breakoff. We will report on the ability of sequential learning to classify a survey as a potential breakoff case during the survey. These predictions will result in reduced breakoffs by allowing for adjustments to engage respondents who are following a sequence terminating in a breakoff.

Predicting Breakoff Using Sequential Machine Learning Methods

Leen-Kiat Soh, *University of Nebraska-Lincoln*

Adam Eck, *University of Nebraska-Lincoln*

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

The increased collection of paradata during computer-based survey administration provides exciting new opportunities for researchers to observe and better understand the processes performed by administrators, interviewers, and respondents (Kreuter 2013). For example, researchers are provided with further context describing respondent behaviors and additional insights into the substantive responses collected by the survey. This is especially true of

AAPOR 70th Annual Conference

response-level paradata, which records the actions of respondents as they complete the survey using the computer system, such as keypresses, mouse clicks, response latency, scroll and navigation behavior, and changes to previously entered responses. However, the additional information contained in paradata is often structurally complex, necessitating sophisticated data processing and analysis techniques. For example, actions recorded in response-level paradata are sequential in time and of variable length: to complete the same survey, one respondent might perform a sequence of 50 actions, whereas another might perform a sequence of over 100 different actions. These properties make the analysis of the relationship between respondent behavior and survey outcomes difficult. Moreover, because respondents will often perform multiple actions per question response, the amount of paradata collected can grow quickly with complex relationships to other data, creating “big data” data sets with additional challenges and concerns. In this presentation, we describe our ongoing efforts to analyze response-level paradata in order to model respondent behavior during survey completion. We are interested in predicting breakoff during self-administered web-based surveys given sequences of respondent actions as part of the Gallup Web Panel. To address the sequential, variable length, and “big data” properties of such paradata, we employ machine learning techniques (sequential classification algorithms) from computer science to predict whether a sequence of actions will ultimately lead to breakoff or survey completion. Preliminary results demonstrate that, with the appropriate algorithm choice, breakoff can be predicted with relatively high accuracy and precision.

Using Paradata for Instrument Evaluation and Refinement

Amanda Reiter, *Mathematica Policy Research*

Sarah Forrestal, *Mathematica Policy Research*

Paradata are a potentially useful tool for improving questionnaire design by helping understand the questions that are more or less effective, the significance of question placement, and the length of time spent on questions by respondents. For a national study of the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), we used paradata to evaluate instrument burden and the order of topic presentation, as well as identify potentially problematic questions. Our innovative use of paradata informed our recommendations for refining the instrument for future administration. We administered a two-part web survey to 90 state-level and approximately 1,800 local WIC agencies (the response rate was 91% to both parts of the survey). In this paper, we will describe how we used paradata to analyze respondent burden, break-offs, and the order in which respondents completed survey modules. Specifically, we will discuss: 1) using question- and module-level burden estimates to identify survey topics to consider revising in future rounds of administration; 2) considering the frequency of break-offs as a marker of problematic questions; 3) examining the number of login attempts as an indicator of how agency respondents shared responsibility for participation; and 4) analyzing the order of module completion to inform their order of presentation to respondents. Using paradata for questionnaire evaluation is a promising approach with the potential to reduce burden and cost and increase data quality in future administrations.

Longitudinal Surveys

How Stable Are Religious Identities?

Becka A. Alper, *Pew Research Center*

Claire Gecewicz, *Pew Research Center*

Jessica Martinez, *Pew Research Center*

Religious identity is a key variable that is widely used by researchers in many different fields. Many surveys include measures of religious identity along with standard demographic measures because it helps us understand a range of social and political behaviors and outcomes. But despite its utility, it is less clear whether religious identity is similar to other identities and characteristics that tend to be stable over a short period of time (e.g., gender, race, educational attainment, etc.) or if it is more like attitudinal variables that can frequently fluctuate. Some respondents may identify with a particular religious group at the time they take a survey because of specific circumstances (e.g., because they attended church earlier that day, or because of the context of the questionnaire), but then identify with another religious group or with no religion only a short time later. Using data from the Pew Research Center's American Trends panel, we examine how common it is for respondents to change their religious identity between two waves of data collection spread out over just a few months. We find that religious identity does in fact vary between our waves of data collection and demonstrate the patterns of movement into and out of religious groups. We assess whether these changes indicate that respondents had weak religious identities to begin with or real religious change (i.e., religious conversion). We examine the religious, social, and demographic characteristics of those who report different religious identities at time 1 and time 2. Our findings suggest that for some, religious identity can be in flux far more often than what research relying on cross-sectional data can indicate. The results of our study underscore the variability of religious identity and we discuss the implication of these findings.

The Effect of Initial Recruitment Efforts on Completion Rates of Subsequent Waves in a Longitudinal Panel Study

Eran N. Ben-Porath, *SSRS*

MollyAnn Brodie, *The Kaiser Family Foundation*

Bianca DiJulio, *The Kaiser Family Foundation*

Survey research often utilizes recruitment efforts such as repeated contact attempts and refusal conversions to increase response rates and reduce possible nonresponse bias. In longitudinal panel studies the efforts to reach reluctant respondents may create a paradox: While multiple contact-attempts and refusal-conversions improve response levels in the initial contact, the reluctant respondents may be less likely to complete follow-up interviews, and, consequently, this would lead to lower completion rates for subsequent waves. This paper assesses the effect of Wave 1 efforts on cooperation in consequent waves, looking at the first three waves of an RDD panel study of uninsured Californians. Beyond the main effect of completion rates for follow-up interviews, we assess the implications of the effect on the study's substantive findings. Our findings are twofold. First, our analysis finds that the number of call attempts during Wave 1 is negatively associated with the propensity to complete Waves 2 and 3. In other words, less effort at Wave 1, would have yielded more interviews in the follow-ups. Second, in assessing the data quality, we find that nonresponse in Wave 2 did not lead to notable bias. That is, the composition of the sample responding to the second wave, when weighted, closely resembled the full group of Wave 1 respondents across a variety of demographic, behavioral and attitudinal

AAPOR 70th Annual Conference

factors. At the same time, there would be no meaningful change in the weighted Wave 1 data, were we to exclude the reluctant respondents, which, in turn may have allowed for a larger Wave 2 sample. These findings indicate a need to consider the utility and risks of recruitment efforts in the context of longitudinal panel studies. Wave 3, which will be completed in the coming months, should provide insight as to the persistence of the effect, eighteen months after the initial recruitment.

Affluent Colleges: Which Students Become Politically Active?

Brianna White, *Wellesley College*

Previous studies have demonstrated that affluent colleges and universities are more likely to be sites of political activism than resource-poor institutions, partially because the students that attend them have been socialized to be active. However, these studies operate under the assumption that all students at wealthy colleges are equally likely to engage in political activism. This study seeks to understand how levels of activism on campus vary between certain social groups (i.e., gender, race, and class). Using the University of California-Los Angeles (UCLA) Higher Education Research Institute, The UCLA survey is a nationally-representative study of college students was administered to college students in their freshman year (1989-1996) and their senior year (1996-1999). This study defines affluent universities as institutions where a majority of students come from a middle-to-upper class background, defined through a family income. In this study political activism is defined through variables such as participation in political groups, voting, protest and petition participation, and service learning. Change in political activism is measured by studying the change between their college and senior year. It is hypothesized that while women experience higher levels of political activism than men because of diversity outreach, lower-income students and non-white students have lower levels of political activism than their counterparts because of barriers to participation. The results of this study can suggest whether affluent campuses are promoting political activism among all groups. This is essential because affluent universities often have the resources and networks to inspire political and social change. If this theory proves to be true, it has important implications for affluent institutions and their promotion of political expression among traditionally under-represented groups (non-whites and lower-income students). Another implication is whether despite educational attainment, certain groups (lower-income and non-whites) are less politically active.

Predicting Panel Attrition on a National Study: How We Can Optimize Locating Resources and Methods

Lauren M. McNamara, *NORC at the University of Chicago*

Ned English, *NORC at the University of Chicago*

Melissa Heim Viox, *NORC at the University of Chicago*

Katie Dekker, *NORC at the University of Chicago*

Ron Hazen, *NORC at the University of Chicago*

National longitudinal studies depend on the retention of diverse populations across an extended period of time. All longitudinal studies run the risk of introducing bias via differential attrition due to movement and nonresponse as the propensity to attrite is not evenly distributed across the population. To gain a better understanding of who is likely to become lost to follow-up, analysis and modeling comparing lost to retained participants were performed using data from a large longitudinal national study of children's health and development. Once women were enrolled in the study, the first participant interview was scheduled to occur before childbirth. Women were then interviewed at childbirth and again at regularly-spaced intervals every three to six months

AAPOR 70th Annual Conference

for the first few years of the study. Women who were 'lost' were searched for using best practice locating methods: contacting non-household contacts, batch locating, interactive locating with free and proprietary databases, and in-person locating. Our paper uses logistic regression to understand panel attrition and predict the propensity of a respondent to become lost to follow-up. In so doing we pursue how attrition is dependent on geography, number and distance of moves, recruitment strategy, demographics, and health-related factors. Our results indicate that it may be most effective to focus resources in the form of extra locating effort or interim follow-ups on a subset of respondents, while decreasing efforts spent on low risk or "stable" respondents. Our findings shed light on the relationship between participant characteristics and panel attrition in longitudinal studies of children.

As We (Still) Like It: Social, Religious and Kid's Activities Remain Americans' Favorite Activities

John Robinson, *University of Maryland*

As valuable as time-diary data have been in documenting declines in women's housework and overall gains in free time, the quality-of-life (QOL) implications for times spent on most daily activities (like eating or traveling) remains largely moot. The collection of subjective data on how people feel while doing various activities provides an important advance in translating time figures into QOL terms, as when people spend more time doing things they enjoy. Subjective data on how diary respondents felt about their daily activities (in general) were collected in the 1975 U.S. national diary data and directly "in real time" in its 1985 replication, and these two aggregate activity rating sources were found to be highly correlated (0.70), despite their different time contexts. Since 2003, this national time-diary series has been replicated in the American Time-Use Survey (ATUS), now conducted continuously by the U.S. Bureau of the Census, also using random telephone interviews with nearly 150,000 Americans reporting on how they spent their time on the previous day. Since 2010, the ATUS began supplementing these accounts with six psychological questions on how these respondents felt as they were engaged in these daily activities. Thus, the ATUS basically provides a continuous national monitor of Americans' everyday QOL in real time. Analysis of these 2010-13 ATUS ratings reveals that they largely replicate the activity enjoyment ratings found in 1975 and 1985. As in these earlier diary studies (and in a national 1986 UK diary survey), ATUS respondents rated social, religious and interactive child activities most positively, and housework and health-problem activities least positively. However, one major important difference in these recent SWB ratings has been the far lower ratings given to paid work activities. While travel activities rated about average, travel to more positive activities were rated more positively and vice versa.

Determinants of Survey Participation

Understanding Dynamics of Consent Requests in Surveys: Consent to Biomarker Data Collection and Administrative Data Linkage in the Health and Retirement Study

Colleen A. McClain, *Michigan Program in Survey Methodology*

Sunhee Lee, *Michigan Program in Survey Methodology*

Jessica Faul, *University of Michigan*

Stefany Barba, *University of Michigan*

Increasing collection of auxiliary information within the survey context—including collection of biomarkers and permission to link to administrative data— has led to significant gains in researchers' ability to draw conclusions from a wealth of data points. However, such designs introduce the possibility of differential bias arising from respondents' noncompliance, similarly to traditional nonresponse bias in survey data. Existing research is important but limited in terms of quantity and scope of examination; thus far, the outcomes of consent requests have typically been linked to respondents' socio-demographic characteristics, health status, and attitudes toward data linkage, as well as to basic demographics of interviewers. Few studies have integrated data illustrating the actual interviewer-respondent interaction—that is, fully considering the dynamics of information comprehension and decision-making involved in requests. We use data from the 2010 Health and Retirement Study to address this gap. In doing so, we first link consent request results for physical measurements, biomarker collection, and Social Security Administration data linkage with respondent characteristics—moving beyond demographics to explore attributes such as personality traits that may be correlated with consent. Additionally, we extend our focus on the interviewer, exploring the impact of interviewer characteristics and interviewer observations in our analyses and incorporating results of behavior coding of audio-recoded interviews to assess the interviewer-respondent interaction. This behavior coding focused on a sample of recordings stratified by race/ethnicity, interview language, and overall response pattern across multiple consent requests, allowing us to quantify characteristics of the consent context (e.g. presence of a third party), auditory characteristics (e.g. vocal tone, speed of speech), perceptions of the interaction (e.g. friendly, frustrated, or enthusiastic interviewers and respondents), and content of the request (e.g. deviations from the interviewer's script, respondent questions). Combining these data sources allows for more complete exploration of consent dynamics in a face-to-face survey context than previous efforts.

Predictors of Nonresponse to Physical Assessments in a Population-Based Survey of Older Adults

Yuan Zhang, *University of Southern California*

Eileen Crimmins, *University of Southern California*

Jennifer Ailshire, *University of Southern California*

Direct assessments of physical performance are increasingly being collected in large, population-based surveys to characterize physical functioning and disability in the older population. However, non-completion of performance assessments among those who consent to take physical assessments is a unique instance of item nonresponse in surveys that threatens our ability to correctly characterize the functioning and disability status of the older population. If the reasons for not completing the tests are related to the respondents' functioning and physical ability, nonresponse may lead to biased estimates of functioning problems in the

AAPOR 70th Annual Conference

older population. Thus, it is important to understand the nonresponse patterns in physical assessments and to adjust for losses where appropriate. Data for this analysis are from the 2006 Health and Retirement Study. We examine missing data and reasons for non-completion of physical performance tasks. Weighted logistic regression models are used to investigate the factors that affect completion. Results indicate that nonresponse to the tests was not random. Nonresponse was more likely for the assessments that require more physical fitness (e.g., balance and gait speed). For noncompletion because of safety concerns, health conditions or inability to complete the tests, self-reports of poor health status and low functioning abilities predict the completion of the physical assessments. Respondents' cooperation levels toward the interview are significantly associated with the completion of tests. When nonresponse resulted from a lack of a suitable location for the test, interviewer observations of the physical environment in which the assessments take place predict the completion of the physical assessment. We develop adjustments to account for factors related to the noncompletion that allow us to reduce bias in estimates of population disability.

Who Would Refuse? An Exploration of Data Quality Trends and Demographic Characteristics on the General Social Survey (GSS)

Jodie Daquilanea, *NORC at the University of Chicago*

Beth Fisher, *NORC at the University of Chicago*

The General Social Survey (GSS) provides a suitable environment in which to examine issues of data quality and missing data points with respect to respondent characteristics. The GSS is a survey sponsored primarily by the National Science Foundation and conducted biennially by NORC to collect data on the attitudes, experiences, and demographic characteristics of residents throughout the United States. The 2010, 2012, and 2014 rounds of the study in particular included a panel sample of respondents who completed the GSS interview in both of the two previous survey years. We will explore the types of characteristics that make respondents more likely to refuse to provide contact information and sensitive data points such as social security numbers, phone numbers, and reports of drug use. Initial results indicate that older respondents are more likely to refuse to provide their social security numbers than younger respondents, and we will expand this analysis to explore other demographic characteristics, such as sex, race, and geographic region. We will explore whether panel respondents become less likely to refuse certain questions after they complete the previous round, and after they complete two previous rounds to determine whether long-term interaction with a study affects the refusal rate. We will also analyze selected data item refusals over the past ten years to uncover trends in the refusal rates of sensitive items. These findings will add to the body of knowledge on the types of respondents who refuse providing sensitive data points and contact information and therefore the types of respondents who would therefore be more difficult to contact in the future for panel interviews due to lack of contact information. Non-response of individual items and of the interview itself both have an inevitable impact on non-response bias, an overall concern of the survey community.

Examining Sexual Orientation, Race/Ethnicity and Interview Language as Correlates of Nonresponse Using Paradata

Sunghee Lee, *University of Michigan*

Colleen McClain, *University of Michigan*

Karen Fredriksen, *University of Washington*

Hyun-Jun Kim, *University of Washington*

Tuba Suzer Gurtekin, *University of Michigan*

Nonresponse remains as an unrelieved challenge for survey researchers, and the effort to understand its mechanism continues to grow. Reflecting an increasingly diverse population, survey in the U.S. have seen adaptations in the operations. In particular, most surveys are now conducted in languages besides English, and some surveys include questions on sexual orientation. These changes, coupled with the collection of paradata, provide unique opportunities to further advancing our understanding of nonresponse and its mechanisms. This paper examines the role of race/ethnicity, interview language and sexual orientation and their interactions on both unit and item nonresponse. Specifically, this study uses 2013 National Health Interview Survey (NHIS), one of the first general population surveys that collect data on sexual orientation. Additionally, NHIS provides paradata on the contact history, allowing us to assess the level of effort required to obtain response, and includes questions on risk behaviors and financial status as well as those cognitively demanding to respond, which often elicit a high level of item nonresponse. We will first create contact and interview effort levels by compositing multiple variables and use them as proxies of unit nonresponse. Focusing on the sample adults, these nonresponse measures will be examined as a function of race, ethnicity, interview language and sexual orientation simultaneously controlling for other characteristics known to be related to nonresponse, such as age. Nonresponse on sensitive items (e.g., income, weight, financial difficulties, tested for HIV) and difficult recall questions (e.g., tetanus shot in the last 10 years, wait time for doctor's appointments) will be examined in a similar way. We will further examine whether or not these characteristics are related to unit and item nonresponse in a similar way.

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

Mini-Conference: Comparing Probability and Non-Probability Samples

Survey Estimation: How Different Are Probability and Non-Probability Survey Designs?

Jill A. Dever, *RTI International*

Bonnie E. Shook-Sa, *RTI International*

Probability sampling designs, those with samples selected with a reproducible random mechanism, are considered by many to be the gold standard for surveys. Theory has existed since the early 1930's to produce population estimates from these samples under labels such as design-based, randomization-based, and model-assisted estimation. This theory ultimately

AAPOR 70th Annual Conference

requires that the sample units, excluded from the analysis files either because of non-sampling or nonresponse, are missing at random. This condition, however, is not always attainable. Studies involving samples without a necessarily reproducible design, referred to as non-probability surveys, have gained more attention in recent years but they are not new. Touted as cheaper, faster (even better) than probability designs, these surveys capture participants through various methods such as respondent-driven sampling or opt-in web surveys. For surveys required to produce population estimates to meet their stated fit for purpose, the link between the sample and the target population as well as the probability of participation must be addressed to justify the desired level of quality. Survey weights or analytic models have been proposed to provide the needed evidence of the data's utility, but research findings on the effectiveness of these approaches are inconsistent. This suggests that probability and non-probability surveys sometimes "work" and sometimes they "fall apart." Through this lens, we argue in this paper that probability and non-probability surveys are not two sides of a research coin but actually lie on a quality continuum. We first review the published material on the definition of quality, keeping in mind that surveys have a specific fit for purpose within this context. Next, we summarize research to date to measure the quality of non-probability survey estimates and compare these criteria with similar probability surveys. We conclude with components for a quality framework that encompasses all surveys to enable their objective comparison.

A Comparison of Surveys Based on Probability Versus Non-Probability Sampling Approaches

Gordon B. Willis, *National Cancer Institute, NIH*

Sadeq R. Chowdhury, *Agency for Healthcare Research and Quality*

Janet S. de Moor, *National Cancer Institute, NIH*

Donatus Ekwueme, *Division of Cancer Prevention and Control, Centers for Disease Control and Prevention*

Erin Kent, *National Cancer Institute, NIH*

Benmei Liu, *National Cancer Institute, NIH*

Steven Machlin, *Agency for Healthcare Research and Quality*

Lisa Mirel, *Agency for Healthcare Research and Quality*

Stephanie Nutt, *LIVESTRONG Foundation*

Juan Rodriguez, *Division of Cancer Prevention and Control, Centers for Disease Control and Prevention*

Anita Soni, *Agency for Healthcare Quality and Research*

Katherine S. Virgo, *Emory University*

Maggie Wilson, *National Cancer Institute, NIH*

Given that response rates for probability-based sample surveys have been dropping precipitously, and because these surveys are generally expensive to conduct, there has been substantial interest in the use of non-probability-based surveys as an alternative. What is unclear, however, is the degree of bias and resultant effects on data quality introduced through use of a non-random selection approach—even when appropriate weights are applied—on both parameter estimates and measures of statistical association. As an attempt to empirically characterize the effects of sampling strategy, the current study compared the results obtained when the same questionnaire items were administered within both probability-based and non-probability-based surveys. As a probability-based survey we selected the 2011 Medical Expenditure Panel Survey (MEPS) Experiences with Cancer survey, a self-administered follow-up to the base MEPS that is representative of the U.S. civilian non-institutionalized population

AAPOR 70th Annual Conference

(with final response rate of 49.4%) – and was either distributed in person or (rarely) mailed to adult cancer survivors. Our non-probability-based survey was the LIVESTRONG Cancer Survivor Study, a web-based, opt-in instrument made available to cancer survivors, and which was not dependent on a statistical sampling approach but instead promoted through email, Facebook, Twitter, newsletter; and professional contact (no response rate could therefore be computed). Our evaluation includes 1,203 and 5,394 adult (18+) cancer survivors from the MEPS and LIVESTRONG surveys, respectively. We will determine the extent to which the LIVESTRONG data, subsequent to the application of several raking and/or post-stratification weighting strategies, match estimates obtained from the MEPS survey, for purposes of population parameter estimation, and for measurement of key associations (through statistical regression analysis). With appropriate caveats, the implications for uses of a web-based survey involving non-probability sampling will be discussed.

A Model-Over-Design Integration Approach in Estimation from Purposive Supplements to Probability Samples

Avinash C. 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 making suitable inferences from purposive samples are again at the forefront because design-based methods are clearly not suitable. There is, however, the possibility of using model-based methods but they require the design to be non-informative; i.e., the model holds for the sample which allows for an unbiased prediction of the unseen from the seen. Also, they are prone to possible model misspecification resulting in bias even if the design is non-informative. An alternative approach termed Model-Over-Design Integration (MODI) for a simplified problem assuming the availability of a purposive sample as a supplement to the probability sample is proposed under the joint design-model randomization. In MODI, a regression model is first fit using the probability sample taking design into account. A design-based estimator; e.g., GREG, is constructed which uses a model-bias correction to the mean predictor (i.e., the synthetic estimator). Next, the above model-bias correction is improved by using another bias estimator from the seen in the purposive sample assuming that the model errors are uncorrelated with the conceptual selection probabilities of units that could be selected in the purposive sample. It can be shown that the MODI method uses shrinkage-type estimation and preserves the desirable property of asymptotic design consistency of the GREG estimator. This may lead to considerable gains in efficiency when dealing with large inexpensive supplements. We remark that while the initial probability sample is used for estimation of model parameters to obtain a synthetic estimator as well as an estimator of model-bias, the purposive supplement is only used to improve the bias correction from the additional seen units. The mean squared error (MSE) of the resulting estimator under the joint design-model randomization can be estimated via replications.

Revisiting Sample Frame and Mode Effects: A Comparison of Point Estimates

Nancy Mathiowetz, *University of Wisconsin-Milwaukee*

J. Michael Brick, *Westat*

Sarah Cho, *SurveyMonkey*

Jon Cohen, *SurveyMonkey*

Kyley McGeeney, *Pew Research Center*

Scott Keeter, *Pew Research Center*

During the summer and fall of 2014, The Pew Research Center, Westat, and Survey Monkey undertook a study to examine the impact of alternative sampling frames and modes of

AAPOR 70th Annual Conference

administration on estimates across a broad range of question items. The study incorporated three distinct study designs: 1. A web-based survey administered to members of a probability based panel (n= 1509) 2. A web-based survey administered to members of a non-probability based panel (n= 5301); and 3. A paper and pencil mail survey sent to a simple random sample from an address-based frame (n=2668). The intent was to determine to what extent designs of current popular interest render similar or different estimates. In addition, several of the questionnaire items replicate questions from the “gold standard” studies such as the American Community Survey, the National Health Interview Survey, and the Current Population Survey, thereby providing a means for assessment among the three survey designs and with respect to a benchmark. We provide initial estimates from the surveys, highlighting both similarities and differences across the three designs. The findings provide comparisons across a broad range of substantive questions, for example questions concerning health, financial insecurity, and political attitudes. Although the findings are presented with respect to specific questions, the underlying research questions focus on the identification of methodological or substantive features related to dissimilarity across modes and frames. The work expands on previous investigations in this area, most notably those of Ansolabehere and Schaffner (2014) and Callegaro, et al (2014).

Fit for Purpose Community Health Surveys: An Experiment in Three Communities

John M. Boyle, *ICF International*

Lew Berman, *ICF International*

James Dayton, *ICF International*

Tala Fakhouri, *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 data systems generate community measures. National health surveys are too small to provide useful community-level data. Conducting parallel community health surveys could substantially augment the benefits of national surveys, and operationalize findings on the local level. National health data (e.g., physical activity, obesity, cancer screening) becomes more meaningful if it can be compared to local data. Local data is more likely to drive local initiatives, but it becomes more compelling when it can be benchmarked against national trends. Non-probability samples represent a potential alternative to traditional probability samples in “fit for purpose” designs which weigh the key components of overall success—including accuracy, timeliness, and cost. Specifically, non-probability web panels can generate usable sample sizes within the resources of local agencies. However, the sample must yield estimates that can be considered accurate and representative. We use population web panels drawn as Census balanced samples to test the utility of non-probability surveys in three communities. The samples will be fielded by email invitations with follow-up invitations to improve the response rate. The target communities were selected for the availability of both ACS and local BRFSS health data. ACS and local BRFSS data will be used to assess the accuracy of non-probability sampling. The questionnaire was designed to parallel a national probability health survey for national comparisons. A total of more than 400 interviews will be completed in each of the three communities. The generalizability of the findings from this experiment is strengthened by the replication in three communities.

Challenges and Applications of Machine Learning, Record Linkage, Administrative and Auxiliary Data in Survey Research

Using Machine Learning Techniques to Predict Respondent Type from A Priori Demographic Information

Douglas Wettlaufer, *University of Nebraska-Lincoln*

Gregory Scott Atkin, *University of Nebraska-Lincoln*

Hariharan Arunachalam, *University of Nebraska-Lincoln*

Adam Eck, *University of Nebraska-Lincoln*

Leen-Kiat Soh, *University of Nebraska-Lincoln*

Robert F Belli, *University of Nebraska-Lincoln*

Respondents can differ in how well they (1) focus on providing answers to survey questions, (2) remember relevant information, and (3) cooperate in providing answers. In this study, these different types of respondents will be assessed. Understanding the most influential factors that determine the type of respondent a person will be and using that information to predict respondent type before the survey begins can enhance adaptive survey design by allowing surveys to be dynamically tailored to the particular type of respondent taking the survey in order to improve response data quality. This work applies machine learning techniques to data describing respondents of the American Time Use Survey (ATUS)—a time diary survey in which respondents report what they did in a 24 hour period—in order to determine which demographic variables are most influential in determining the type of respondent a person will be as well as to construct models that can predict respondent type before the survey takes place. The respondent types are measured as (1) focused, by the amount of time taken to respond to interviewer questions, (2) forgetful, by the number of “don’t know/can’t remember” responses, and (3) cooperative, by the number of refusals to respond. To predict respondent type, our models use respondent demographic data collected in the Current Population Survey (CPS). The CPS data is collected two to five months before the ATUS interview is conducted, enabling it to be used to predict respondent type prior to the survey taking place. Applying a data-centric, machine learning approach, this work will provide insights into designing survey systems that are more context aware of who the respondent is and enable researchers to collect high quality data from all respondents. We report on the results and discuss the next steps.

I Know What You Did Next: Predicting Respondent’s Next Activity Using Machine Learning

Hariharan Arunachalam, *University of Nebraska-Lincoln*

Gregory Scott Atkin, *University of Nebraska-Lincoln*

Adam Eck, *University of Nebraska-Lincoln*

Doug Wettlaufer, *University of Nebraska-Lincoln*

Leen-Kiat Soh, *University of Nebraska-Lincoln*

Robert F. Belli, *University of Nebraska-Lincoln*

The American Time Use Survey (ATUS) is an annual telephone survey conducted by the Bureau of Labor Statistics in which sample populations of Americans are interviewed regarding how they spent 24 hours from 0400AM the previous day. During the interview process, the interviewers have to (1) maintain a conversation with the respondent to extract the required information, and (2) record the information as reported by the respondent. More often than not, many of the data points that the interviewer records may seem to be intuitively obvious before it

AAPOR 70th Annual Conference

is reported by the respondent based on what was just reported or when it was reported. If such intuitive predictions of the respondent's next activity could be provided by the instrument itself in a non-intrusive and targeted manner, one could potentially use it to speed up the recording process by the interviewer with caution, so as to not bias the interviewer to accept or encourage only certain respondent answers. This work uses probabilistic model based prediction techniques from machine learning to (1) model the data from previously collected and publicly available ATUS data (2010 through 2013) and (2) make next activity predictions from these models. We use the activity responses and the Current Population Survey (CPS) data to develop our models. The CPS data contains 374 variables that provide demographic information about the respondent (collected 2-4 months prior), which enables us to identify routines of different types of people and generate context-aware and targeted predictions. We investigate (1) the effectiveness of the predicted activity from the built models, and (2) the usefulness of these predictions for the interviewer. The predictions are made based on the respondent's previous activity and popular activities at certain times of the day. We examine and report on the investigations and discuss extending these techniques for future work.

Privacy, Data Linkage and Informed Consent

Frauke Kreuter, *University of Maryland, JPSM; IAB*

Joe Sakhaug, *University of Mannheim*

Alexandra Schmucker, *IAB*

Eleanor Singer, *University of Michigan*

Mick Couper, *University of Michigan*

Across agencies within the U.S. and across countries around the world, we see increasing pressure to use administrative data in addition to survey data. Whether it is the use of administrative data as sampling frame, to augment survey data or to replace surveys entirely, in many settings informed consent is required before the data can be used. However, as the research on informed consent to the survey request has shown us, the nature and wording of the consent question can affect consent rates (and presumably non-consent bias). This paper will present result from a large telephone study in German that included a couple of randomized experiments around the consent question. Drawing on findings in behavior economics, consent to pass addresses on to data collection agency was asked with opt-in vs. opt-out wording, and consent to data linkage was randomly framed as gain or loss. Our findings support the striking difference of opt-in opt-out findings found elsewhere, but show interesting interaction effects with gender and age that might shed some light on the nonresponse literature in general. To our surprise the gain gain-loss framing did not lead to notable differences in consent rates. In addition to the experimental findings we will present results from follow-up questions that were issued to examine the understanding of the linkage consent questions.

Auxiliary Data as Correlates of Survey Nonresponse: An Examination for Landline and Cell Phone Sampling Frames

Rajesh Srinivasan, *Gallup*

Manas Chattopadhyay, *Gallup*

Jenny Marlar, *Gallup*

Auxiliary data is often available for purchase from commercial survey sample providers. This information, if correlated with both the survey variable of interest and response propensity, can be valuable for nonresponse weighting adjustments. Yet, the auxiliary data available from commercial vendors often differs between landline and cell phone sampling frames. In an era of

AAPOR 70th Annual Conference

growing dual-frame surveys prompted by a continued erosion of landline telephone coverage, a better understanding of the auxiliary data available for both telephone frames, and its relationship with nonresponse is needed. Using data from the Gallup daily tracking poll, a probability-based national survey conducted nightly by the Gallup Organization, along with auxiliary information from a commercial sample provider, this analysis will investigate the correlates of survey nonresponse. Specifically, we examine the auxiliary data available for landline and cell phone sample frames, and their relationship with survey nonresponse.

A Meeting of Data: Linking Data Across Survey Sources to Estimate Improper Payments of Housing Rental Assistance

Davia Spado, *ICF International*

Kelly Martin, *ICF International*

Sup Thanasombat, *ICF International*

Sophia Zanakos, *ICF International*

Linking data from multiple sources, and in particular from administrative sources, offers a number of benefits to survey statistics—a larger set of analytic variables; the ability to evaluate components of total survey error such as response error and nonresponse bias; and minimized burden to respondents. The Quality Control for Rental Assistance Subsidy Determinations (HUDQC) study, sponsored by the U.S. Department of Housing and Urban Development (HUD), combines multiple sources of data for the analytic purpose of estimating national improper payments of rental assistance, such as incorrect rent determination, among assisted housing programs. The HUDQC model integrates on-site record abstractions obtained from assisted housing administrators, computer-assisted personal interviews (CAPI) with tenants, requested third-party documents, and administrative data from the Social Security Administration. Specifically, income, asset, and expense items from these sources are linked and compared to determine the quality control (QC) value for each item. This presentation will outline the mechanisms of record linking across survey data sources, including administrative data, the challenges encountered during this record linking, and the benefits of this approach for the HUDQC study. Challenges discussed will include evaluating the completeness and accuracy of data from each source, and the selection and validity of key variables used for record linking. Benefits discussed will include discovery of additional data points, verification of the QC items, and cost-savings through use of established administrative datasets. Attendees of this presentation will gain insight into the real-world application of record linking across administrative data and data obtained from other sources.

Perspectives on Polling

How Would We Measure Public Opinion If We Didn't Have Public Opinion Polls? (And Would We Be Better or Worse Off?)

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

Consider the alternative reality in which one wanted to measure public opinion, but public opinion polls did not exist. How would we assess public opinion? We consider three examples of what had, might, or could be done. First, how was public opinion measured by scholars, journalists, politicians, and others in the United States before the advent of public opinion polls in the 1930s? There were many alternative methods being used that were soon eclipsed by and then largely abandoned following the rise of public opinion polls. Second, even in the public

AAPOR 70th Annual Conference

opinion polling era, alternatives have been advanced as either substitutes for or supplements to public opinion polls. A case study of assessing public preference of constitutional changes in Hong Kong in 2014 illustrates the possibilities. Finally, now in the alleged post-public opinion polling period, in the era of the Internet, social media, and Big Data, new alternatives to public opinion polls are being advocated ranging from automatic classification of tweets, to tallying Facebook likes, to counting Google searches and beyond. The comparative advantages and disadvantage of public opinion polls and their alternatives are discussed.

Explaining Variations in Election Surveys: Identifying Contest, Year and Election Type Trends by Combining GAM and HLM Models

Josh Pasek, *University of Michigan*

Election surveys are one of the most important tools we use to follow and forecast electoral campaigns. Yet, the accuracy of these surveys varies widely depending on the type of election we are examining and a variety of methodological choices that pollsters make. Recently, interest has grown in attempts to pool election surveys both within and across contests in order to understand the dynamics at play and to improve estimates of election day behavior by mitigating sources of bias. To test the appropriateness of pooling assumptions, this study models and empirically assesses the similarity of trends over time in different electoral contests as well as across different states for national elections. Using generalized additive mixed models with data from 2,576 statewide horserace polls conducted from 2008-2014, I show that some collections of contests display remarkably similar trends over an election cycle. These trends, however, are clustered by election year, type of office, state, and the incumbency status of a race. An aggregated approach reveals that polling in presidential campaigns is considerably more sensitive to campaign events than polling in senatorial and gubernatorial campaigns and that the 2014 election was notable for the consistency of polling trends across all sorts of contests. In addition to aiding our interpretation of election dynamics, this sort of modeling paves the way for a more nuanced assessment of bias in the results of individual surveys.

Polling the Pollsters: A Survey of State Polling Organization Directors

Kenneth Fernandez, *Elon University*

Jason Husser, *Elon University*

Maggie MacDonald, *Elon University*

In 1985 Wright, Erikson, and McIver described the study of public opinion at the state level as in its infancy. Almost three decades later we find that the field of state public opinion research has grown dramatically. A survey conducted in 2005 by Parry, Kisida, and Langley (2008) found a vibrant community of 54 “active, regular polls of politics and policy in 35 states.” Although no longer in its infancy, state public opinion polling continues to have its share of challenges and barriers. State polls are more likely to be conducted during an election year and more likely to be conducted in a state with a competitive state-wide race. The 2014 midterm election uncovered some neglected areas in the U.S. regarding frequent public opinion polling data. The two-thirds of the states who didn’t have U.S. Senate race were mostly ignored by pollsters, while many of the 34 states that had a Senate race only had one or two state-wide surveys conducted. Although North Carolina had 21 state-wide polls conducted in 2014, most researchers won’t have access to this data because most of the organizations, including academic organizations, don’t readily make their raw data available to the public. In this research project we replicate the survey conducted Parry, Kisida, and Langley (2008) by using an online survey of polling directors of organizations that conduct state-wide surveys. The primary purpose of the survey is the capture the health, growth and diversity of the polling

AAPOR 70th Annual Conference

organizations across the 50 states. Survey items will mostly focus on methodology, but will also ask directors of the challenges facing survey research of state populations.

New Entrances for Exit Polls: Comparing Voter Response in Pre-Election Online Surveys to Traditional Exit Polling

Jennifer J. Agiesta, *The Associated Press*

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

Annie Weber, *GfK Custom Research, LLC*

Emily Swanson, *The Associated Press*

Beyond helping news organizations make election projections, exit polls also provide insight into voters' political attitudes. Typically, voting locations are sampled using a design that aims to provide representative statewide results. Exit pollsters then conduct in-person interviews among voters leaving each selected polling place. Several issues, including the rise of early voting and declining participation in exit polls, have increasingly affected exit polls' accuracy. Larger samples covering more voting places or in-depth sampling of early or absentee voters by phone becomes cost prohibitive. We were interested in looking at results obtained using an online alternative. This study will analyze the results in two states (Georgia and Illinois) of a pre-election, web-based survey of self-identified voters and likely voters drawn from two different sample sources: a probability-based panel (GfK's KnowledgePanel®) and an opt-in panel source. We compare those results (among about 4800 participants) with the National Election Pool exit polls for those states, conducted via in-person and telephone interviews. Our goal was to determine whether voters interviewed in the final days of the campaign express different opinions than those voters included in traditional exit polls. By using nearly identical questions about voting and political attitudes, the two sets of surveys allow for a side-by-side comparison of voter attitudes among Election Day voters who completed their questionnaires online at home compared with those who filled out a paper questionnaire as they left their polling place, as well as comparing pre-election results collected among early or absentee voters by phone for the NEP or online for the Associated Press and GfK. We will report on attitudes among key demographic groups within the voting population identified for each sample, and a poll-by-poll comparisons of the makeup of those who backed each candidate in the top race in each state.

2014: Right Turn on a Bumpy Road

Gary Langer, *Langer Research Associates*

Damla Ergun, *Langer Research Associates*

Julie Phelan, *Langer Research Associates*

Gregory Holyk, *Langer Research Associates*

Some observers called it "the election about nothing" but the 2014 midterms in fact were about very much – chiefly, an extensive, long-running economic and political discontent that's been reshaping party loyalties and upending national politics for the last decade. Using a combination of our pre-election polling for ABC News in the ABC News/Washington Post poll, historical data from past years and exit poll results, we'll present our analysis of how 2014 voters came to their choices. Our presentation will include correlational analyses of economic and political attitudes with election outcomes, as well as a regression analysis evaluating the strongest predictors of the 2014 vote.

Cross-National Research on Income and Economy

Legal Authority and Crime Control: A Comparative, Cross-National Analysis

Jonathan Jackson, *London School of Economics & Political Science*

Mike Hough, *Birkbeck College, London*

Ben Bradford, *University of Oxford*

Jouni Kuha, *London School of Economics & Political Science*

Why do people obey the law and cooperate with legal authorities? Two theories of human motivation and legal regulation typically dominate public policy and debate. An instrumental account focuses on deterrent threat and coerced compliance with the law. It posits that people are responsive primarily to the risk of punishment in deciding whether or not to break the law. If this is true, agents of criminal justice need to send out signals of effectiveness, strength, deterrence and detection. Social control mechanisms and credible risks of sanction would persuade rational-choice individuals that - while otherwise desirable - a criminal act is not worth the risk. Normative theories begin, by contrast, with the premise that most people comply with the law and cooperate with legal authorities largely out of ethical or moral considerations. When the police and criminal courts wield their authority in fair and neutral ways, this gives them legitimacy in the eyes of citizens. Legitimacy then shapes behaviour because people internalise the value that it is right and proper to obey the law and cooperate with legal authorities. To test these two theories – both of which place public opinion centre-stage – we analyse data based on national probability samples of 30 countries. We were involved in the design of a rotating module in Round 5 of the European Social Survey, covering 27 countries. We also replicated the module in new probability sample surveys of adults in the United States, South Africa and Japan. Among other things, we measured people's trust in legal authorities and people's beliefs about the legitimacy of the police and the law. The resulting dataset provides the first cross-national comparative assessment of legal compliance, covering diverse social, legal and political contexts. We report some of the key findings from the project.

Assessing Tax Compliance in Europe Through Public Opinion: Quantitative Methods and Qualitative Considerations

Joseph M. Goldman, *The Gallup Organization*

Given the current fiscal challenges of many European states, research on tax compliance within Europe has received a great deal of attention from policy makers. Assessing the reasons behind tax evasion is a difficult task from both survey research and econometric perspectives. This paper focuses on measuring and interpreting the intrinsic motivation to pay tax, or "tax morale." Building on previous research, my model incorporates the multi-level methodology of Lago-Peñas and Lago-Peñas (2009) and the ground-breaking connections between tax morale and tax compliance by Halla (2010). My model compares factors of tax morale from Torgler (2004) with modified "underground economy" estimates provided by Schneider (2011). Much of the tax research community's survey data, collected through the European Values Study, offer a significantly different language style than other tools, such as the Gallup World Poll, provide. These different tax assessment sources are discussed in terms of how to best understand the root cause of tax evasion. My methodology utilizes individual perceptions on a variety of social and cultural issues measured by the European Values Study and the Gallup World Poll and compares them to national-level tax compliance data. The results of my study indicate that Europeans who have confidence in their social security systems and respect the laws of their country are more likely to reside in tax-compliant locales. However, the more nuanced wording

AAPOR 70th Annual Conference

and limited response options of Gallup World Poll questions might encourage a more accurate depiction of tax morale in Europe. Given the higher consistency with national tax evasion data of more flexible question language and less arbitrary quantitative analysis, new methods of analysis may lead to a better understanding of why Europeans pay tax and how governments can best respond.

Closing Data Gaps in Global Employment Metrics

Benjamin Ryan, *Gallup*

This paper aims to assess the ability of third party nationally representative surveys with modest n-sizes to validate and, where necessary, fill gaps in country level economic metrics around the world, especially in regions with infrequent, inferior, or nonexistent official reporting. Global economic metrics have long been essential tools for international businessmen and policy makers for some time now. However, the quality and consistency of official data varies significantly from country to country and region to region. What's more, data may not be reported on a regular basis; some countries, especially outside of Europe or the Americas, may report only once or twice per decade. I explore these concerns through the lens of global employment figures collected by the International Labor Organization (ILO) in the ILOSTAT database. I begin with a review of the relevant sources, definitions and methodologies. I then compare these employment figures with the corresponding Gallup World Poll figures on employment status for all available countries in each of the years 2010-2014, identifying areas of convergence and divergence between the datasets. Overall, I find a high degree of agreement between Gallup's employment figures and the available official national reporting in corresponding countries and years. Additionally, Gallup country-level figures with no corresponding official reporting that year generally remain consistent with preceding and subsequent official reporting, indicating a good fit for filling data gaps. Interesting divergences are seen in Sub-Saharan Africa and Latin America, where Gallup more frequently finds substantially lower employment levels, and in Eastern Europe, where Gallup more frequently finds substantially higher employment levels, compared with ILO-curated statistics. I discuss potential causes and remedies for these divergences, including sampling, survey mode, respondent or household selection, definitions, seasonality, and other factors. I conclude by discussing potential data validation and integration strategies for improving worldwide employment metrics.

Assessment of Community Basics: Evaluating Country Development Using Local Citizen Ratings

Cynthia English, *Gallup*

Elizabeth Steele, *Gallup*

Development experts rely heavily on outcome indicators, such as GDP growth, literacy rates, and number of paved roads, to evaluate a country's economic development. However, a key piece of data is often missing from these evaluations: perceptions of the people who rely on the underlying institutions and infrastructures responsible for these outcomes. Nationally-representative surveys conducted annually by Gallup ask residents in more than 160 countries to assess the quality of local infrastructure, including transportation, education, healthcare, housing, and the environment, among other topics. Together these measures are aggregated at the country-level in the Community Basics Index (CBI), which provides a unique, reliable, and comprehensive assessment of local infrastructure through the perceptions of residents. This paper will explore the relationship between public opinion about the accessibility and quality of local infrastructure, when aggregated at the country-level, and macro-level development

AAPOR 70th Annual Conference

indicators, such as the United Nations Human Development Index which combines outcome-based health and education metrics with gross national income. One assumption under investigation is that countries with higher levels of development will be more likely to have quality community services and infrastructure, and that this quality will translate into higher levels of satisfaction among residents in these countries compared to those in less-developed countries. This theory will be tested and outlier countries where public sentiment about community basics is either much more positive or negative than other countries with similar development will be discussed. While a modest correlation between country-level CBI scores and select development indicators is expected, the paper will also explore the relationship between individual measures contained within the CBI and indicators specific to that sector, such as satisfaction with education system and literacy rates. Public opinion data, alongside important outcome-based development indicators, provide invaluable information for institutions working to improve experiences and outcomes for people around the world.

Estimating the Joint Effect of the Corporate Hegemony in Labor Relations and Democratic Public Perception on Income Inequality

Soyon Kim, *SUNY-Stony Brook University*

To explain cross-national variations in income inequality, this study is designed to disentangle the net effect of the perception of economic democracy on income inequality, given the corporate hegemony supported by labor market deregulation of the country. The perception of economic democracy is captured as the public opinions about income equality, taxation, unemployment aid and equal job opportunity for women. Corporate hegemony in labor relation is two-dimensional: legal arrangements for labor deregulation of each country, and public confidence about major companies. For the empirical analysis, the data are a compilation of the three datasets from the Economic Freedom of the World index (EFW), the World Values Survey (WVS) and the UNU-WIDER World Income Inequality database (WII). These provide indicators of labour deregulation, public perceptions and gini coefficients, respectively. The WVS data are aggregated to the national level. I selected the high-income OECD countries (1990- 2010). A two-way fixed effects regression model is used to control the unmeasured time-invariant variables that are omitted from the regression model. Then, the model includes an interaction effect of democratic perception and political action of each country on income inequality, which is expected to be conditional upon the corporate hegemony. The preliminary finding from a simple correlation analysis shows that labor deregulations have a positive strong association with income inequality and so do the public perceptions that are less congruent with economic democracy. The full results are expected to support the argument that neoliberal labor deregulation offers corporate autonomy that upholds the hegemonic position of major companies of a society, so that the less protected labour rights are associated with the higher level of income inequality.

Tradeoffs Between Response and Accuracy

Eliciting Financial Information on Surveys: The Tradeoff Between Precision and Nonresponse

Brooke Helppie McFall, *University of Michigan, Survey Research Center*

Michael Gideon, *University of Chicago*

Joanne W. Hsu, *Federal Reserve Board of Governors*

Precise and accurate measures of household financial characteristics are critical for empirical examinations of key economic behavior and relationships, like consumption, borrowing, and retirement decisions. Collecting financial data on surveys is difficult due to the perceived sensitivity of personal finances and the cognitive burden on the respondent. Both issues may result in satisficing behaviors, such as rounding, claiming not to have some categories of assets, or item non-response, that reduce the quality or quantity of data collected. Major economic surveys address these challenges by asking first for “exact” values, then allowing range responses as an alternative. This allows respondents unwilling to report an exact value to provide some potentially useful data for researchers, especially relative to item nonresponse. However, exact value responses provide more information than ranges only if they are more precise than ranges. If exact value responses are rounded numbers no more precise than range options, there is little marginal information from eliciting values. This paper explores the prevalence of rounding, exact values, and item nonresponse on income questions on three surveys with different range alternatives. We analyze the determinants of rounding behavior, particularly the role of education, wealth, cognition/financial literacy, and survey mode. We assess how estimates of key economic relationships, like savings rates, are influenced by rounding. Our results inform questionnaire design by describing the tradeoff between precision and nonresponse when eliciting exact values versus ranges. The determinants of rounding behavior, particularly how systematic it is, has implications for economic research using survey data, particularly on financial well-being and income inequality. Preliminary evidence shows that respondents tend to round to similar significant digits regardless of income levels, which suggests that those with high values of income will introduce disproportionate error into sample means, and panel measurements of income changes may be biased.

Item Nonresponse in a Mixed-Mode Household Travel Survey

Reyna J. Peña, *Abt SRBI*

Melanie Wilbur, *Abt SRBI*

Stas Kolenikov, *Abt SRBI*

Item nonresponse is a common indicator of data quality. Demands for data quality are frequently balanced against greater project efficiencies and cost savings considerations in mixed-mode household surveys. To the extent item nonresponse can be relied upon to evaluate data quality, developing strategies for item nonresponse minimization will be desirable. Early research on item nonresponse in surveys leaned on cognitive and social psychological frameworks to explain effects associated with face-to-face interviewing, respondent burden, motivation to comply, topic salience, perceived threats, and privacy needs. Technological advances in data collection have obviated the need for some item nonresponse minimization practices while also introducing entirely new challenges. Drawing on data from a recently completed household travel survey conducted in the MidSouth region of the United States, we uncover both common and unique patterns of item nonresponse. In this presentation, we analyze effects on item nonresponse from: a) data collection mode; b) question content type; c)

AAPOR 70th Annual Conference

differential incentives; d) household characteristics; and, e) proxy reporting procedures in household travel surveys. Preliminary findings suggest: a) data quality concerns were greater for interviews completed exclusively on the web compared to mixed web+paper/phone+paper/phone+web modes; b) item nonresponse rates were higher for personal work/school location reporting in comparison to other destination reporting; c) promised incentive structure yielded mixed benefits; d) respondents who used initials rather than full household member names had significantly higher item nonresponse rates to personal questions like income, household member age, education, place of employment; and, e) self-reporting was more reliable than proxy reporting overall. We discuss strategies to reduce item nonresponse including obtaining complete recontact information from respondents, assisting the primary respondent in garnering cooperation among all household members, and applying questionnaire design techniques aimed at lessening respondent burden and increasing trust in sponsorship.

Estimating Change in Telephone Survey Bias in an Era of Declining Response Rates and Transition to Wireless Telephones - Evidence from the National Immunization Survey (NIS), 1995-2013

James A. Singleton, *Centers for Disease Control and Prevention*

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

Meena Khare, *Centers for Disease Control and Prevention*

Holly Hill, *Centers for Disease Control and Prevention*

David Yankey, *Centers for Disease Control and Prevention*

Vicki Pineau, *NORC at the University of Chicago*

Kirk Wolter, *NORC at the University of Chicago*

During the past twenty years, telephone survey response rates have steadily declined and cell telephones have replaced landline telephones as the main source of telecommunication. The CASRO response rate for the National Immunization Survey (NIS) landline frame declined from 87.1% in 1995 to 62.3% in 2013. Starting in 2011, the NIS used a dual sampling frame, adding cell telephone numbers to the landline sample. These changes underscore the importance of assessing bias in survey estimates over time and adapting survey methods. A previous evaluation of the NIS 1995-2004 with landline RDD samples showed minimal bias in survey estimates. The purpose of this paper is to assess changes in bias of vaccination coverage estimates from the NIS 1995-2013. NIS is a large RDD telephone survey initiated in 1994 to annually assess vaccination coverage among children 19-35 months, and consists of a household telephone survey followed by a mail survey of vaccination providers to collect the child's vaccination history. To assess changes in bias in selected vaccination coverage estimates from the NIS, we use data for children with overlapping birth cohorts in successive NIS annual samples (e.g., the 2011 and 2012 NIS samples both included children born January 2009 – May 2010). Trends in annual bias are assessed by comparing these consecutive bridging cohorts, using a consistent metric not subject to recall bias (i.e., using provider-reported vaccination histories of children as of age 19 months). Change in bias is estimated by controlling for factors typically used in NIS weighting adjustments. Results are compared to external vaccination coverage estimates from the National Health Interview Survey samples to estimate actual bias in annual estimates. This approach can provide an ongoing assessment of bias in NIS estimates and evaluation of new approaches (e.g. drop landline frame or switch to address-based sampling frame).

Response Rates vs. Representative Data: Is the Oversampling of Listed Sample on the BRFSS Survey Helping Response While Reducing Data Quality?

Piper Jean DuBray, *ICF International*

Randy ZuWallack, *ICF International*

Kristie Healey, *ICF International*

The Behavioral Risk Factor Surveillance System (BRFSS) is a dual-frame RDD telephone survey sponsored by the Centers for Disease Control and Prevention. The landline RDD sample is disproportionately stratified based on a list-assisted frame with listed telephone numbers oversampled relative to unlisted numbers at a rate of 1.5: 1. The oversampling of listed phone numbers provides efficiency and increased response. However, analysis of BRFSS landline data have shown the unlisted sample tends to include a younger population (under 35: 5% listed, 11% unlisted), lower income (under \$25,000: 26% listed, 33% unlisted) and a higher percentage of minority respondents (Hispanic: 5% listed, 10% unlisted.) While response will be higher when oversampling listed numbers, the distribution will skew toward an older, more affluent, less diverse sample of respondents. We evaluate alternative sampling strategies with scenarios that include proportional allocation as well as oversampling unlisted numbers. The simulated landline results will be combined with actual cell phone results. Our evaluation will focus on the quality of sample distributions using a representative indicator as well as the impact on cost and effective sample size.

Innovations in Nonresponse Bias Measurement and Reporting for Probability-Based Web Surveys

Martin Barron, *NORC at the University of Chicago*

Ned English, *NORC at the University of Chicago*

Paul Lavrakas, *NORC at the University of Chicago*

While survey researchers have long taken a holistic view of survey quality through the total survey error perspective, this is typically not how we discuss or summarize our survey quality. Thus, as survey researchers have argued the limits associated with reporting the response rate as a single gauge of survey quality, it has remained the most widely used, cited, and publicly salient measure of a survey's quality and trustworthiness. A number of alternatives have been proposed—such as representativity indicators and realization rate—yet these alternative rates have not been widely adopted nor accepted as alternatives to the response rate. For web surveys sourced by a probability sampling, the AAPOR response rate tends to be low. Consequently, web surveys based on probability sampling have historically had some difficulty in achieving respectability in the field. There tend to be lingering concerns about the self-selection and non-response bias potentially undermining the validity and reliability of web surveys based on probability sampling. In order to provide our clients with a fuller and more complete summary of the quality of our web survey samples and survey results, NORC is developing a standard sample quality report that will accompany NORC's surveys. This report, in addition to presenting information required by the AAPOR transparency initiative, will provide several different measures of survey measures of representativeness and non-response bias. In our paper, we describe the innovations in commercial databases matched to address samples and in analytic techniques that make possible the quantitative assessment of sample quality for web surveys.

Using Propensity Scores, Sequencing Interventions, Subsampling and Other Strategies to Implement Responsive Designs

Adaptive Design in an Establishment Survey: Applying Data Collection Procedures Strategically in the Agricultural Resource Management Survey

Jaki S. McCarthy, *USDA/ National Agricultural Statistics Service*

Tyler Wilson, *USDA/National Agricultural Statistics Service*

In household surveys, adaptive design has been applied to strategically allocate data collection resources to reduce costs or improve data quality. For example, rescheduling or reprioritizing calls during data collection may improve response rates for lagging subgroups within a sample. Adaptive design strategies can be applied in establishment surveys, and have unique differences from most household surveys. Establishment surveys may have richer frame or auxiliary data available, including historical response information from prior data collections. This may allow for more pre-planning of optimal data collection strategies rather than relying solely on responsive strategies. Another critical feature of many establishment surveys is that the survey population may be small or there may be a small number of high impact sample units that represent a major share of a given survey estimate. Nonresponse from these units has a much larger potential impact on estimates than nonresponse from smaller units (even though these smaller units may have much larger weights). This talk will discuss the unique considerations for adaptive design in establishment surveys and recent research on applying adaptive design concepts in NASS's Agricultural Resource Management Survey. For this survey, different data collection strategies were applied to three sets of sample units: those with the highest nonresponse propensities, but critical to nonresponse weighting were assigned to field office directors rather than interviewers for recruitment; those with midrange likelihood of nonresponse were assigned to interviewers for in-person recruitment and data collection; and the remaining units were mailed questionnaires with interviewer follow up. The presentation will discuss the rationale behind these strategies and research exploring how these strategies compare to the traditional data collection procedures applied uniformly across the sample.

Evaluating Sequence of Responsive Design Interventions on the Health and Retirement Study Using the Sequential Multiple Assignment Randomized Trial (SMART)

Piotr Dworak, *University of Michigan*

Inbal Nahum-Shani, *University of Michigan*

Wen Chang, *University of Michigan*

Responsive survey design typically includes sequences of case-level interventions aiming to minimize survey non-response. Effectiveness of individual interventions (such as special letters, incremental incentives, and other persuasion strategies) in minimizing non-response has been traditionally evaluated in experimental research designs focusing on comparing across intervention alternatives. However, such designs cannot tell us how to best sequence and adapt interventions (or their combinations) over time to maximize data collection outcomes (response rate, cost, time, and bias). An adaptive approach to sequencing interventions is needed given that some types of interventions may be more suitable for different non-respondent subgroups. Moreover, interventions vary in terms of cost and a cost-effective approach to responsive survey design should seek to sequence interventions beginning with the less costly strategies and reserving more costly strategies to those failing to respond repeatedly. Lastly, alternative sequences of interventions may have different impact on response rate, data collection time, but

AAPOR 70th Annual Conference

also non-response bias. The current research applies the Sequential Multiple Assignment Randomized Trial (SMART) to optimize the responsive design features of the Health and Retirement Study and answer the following research questions: What is the optimal initial intervention for non-respondents (increase incentive or continue with traditional persuasion techniques)? What are the best subsequent strategies for continued non-response following the initial intervention (continue with the initial intervention, intensify the initial intervention, or switch to the alternative)? Since this is the first implementation of SMART -- a research design developed for optimizing adaptive health interventions -- in survey methodology, this research focuses mainly on the feasibility and applicability of SMART design approach, but also on the primary outcomes such as response rates, reduction of data collection time, and cost-effectiveness.

The Use of a Calibration Sample in a Responsive Survey Design

David Wilson, *RTI, International*

Jennifer Wine, *RTI, International*

Bryan Shepherd

Efforts to decrease nonresponse bias through responsive survey designs commonly use targeted interventions to increase response rates for underrepresented groups. The success of such approaches depends on two necessary conditions: effective interventions and a correlation between response propensity and the survey variable. Both of these conditions can vary with time, meaning that survey design decisions based on old data may not provide an optimal research design for a current data collection. The more similar the sample used to evaluate the impact of proposed interventions is to that of the current data collection, the lower the risk of time-based effects. Furthermore, for many studies, prior data on which preliminary assumptions can be based may not even exist. The design of the 2012/14 Beginning Postsecondary Students Longitudinal Study (BPS:12/14) included the use of a subsample, selected from the primary sample and fielded just weeks ahead of the remaining cases, that allowed researchers to evaluate the impact of proposed data collection protocols for the entire sample. As a result of this design, researchers evaluated multiple incentive options and used them to optimize the data collection for the majority of respondents, with minimal risk of time-lag effects. Additionally, the exact data collection procedures, equipment, and protocols to be used for the larger data collection were tested, providing an early warning system for errors. In this presentation, we review the BPS:12/14 responsive design approach with an emphasis on the role of the subsample in providing information that was used to adapt the survey protocol for the majority of cases. We discuss the advantages and disadvantages of this approach and provide an example of a bias-reduction effort that was experimentally evaluated in the subsample and informed decisions for the larger group of sample members.

Exploring the Impact of Design Choices on Propensity Model Performance and Stability

James Lawrence, *U.S. Census Bureau*

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

Chandra Erdman, *U.S. Census Bureau*

Travis Pape, *U.S. Census Bureau*

Given declining response rates, increasing costs, and other resource constraints, there is a clear need to have a good understanding of expected response rate variation that may affect decisions on data collection effort. Understanding this variation can be useful to establish when additional field effort on a case may not be cost-effective and to determine which cases may be

AAPOR 70th Annual Conference

closer to completion. This paper presents the statistical techniques used by the Census Bureau to predict response propensities for the next attempt on a case, given its known characteristics as well as the results and feedback from prior attempts on that case. Known characteristics of the case before the start of data collection, including block group-level data and sample frame data, and survey paradata obtained from prior contact attempts during data collection were used in this analysis to model attempt-level response rates. We will describe the steps to identify the best predictors of response rate, model design choices that affected the ability to predict response rates, how model and variable choices affect the stability of predicted propensity distributions from month to month, and how much predictive power the best models have. These attempt-level propensities, updated for each open case every night, will be used to inform the field data collection. We will discuss the reliability of the propensity estimates, and how we address cases with insufficient information to predict. The creation of daily response propensities for use in managing field work may be a helpful tool in promoting efficient data collection. There are many challenges to designing a good propensity model.

Coverage and Survey Participation

The Common Cause Model, Coverage, and Key Outcomes from the National Crime Victimization Survey

Rachel Michelle Bray, *University of Maryland Joint Program in Survey Methodology;*
U.S. Census Bureau

The common cause model is often used to describe the mechanism underlying the effect of survey nonresponse on survey outcomes. Under the common cause model for nonresponse, the response propensity and outcome of interest are related through a third underlying factor. This paper establishes that a common cause model may also be used to describe the relationship between coverage and survey outcomes. Under the common cause model for coverage, the coverage propensity and outcome of interest are related through an underlying factor. This paper demonstrates how the common cause model for coverage could be reasonably applied to describe the mechanism underlying coverage and key outcomes in the National Crime Victimization Survey (NCVS). First, a set of characteristics associated with coverage of the NCVS frames is identified through a literature review. Then, logistic regression is used to establish whether the coverage characteristics are also predictive of survey outcomes. The results show that coverage characteristics including tenure, housing unit structure type, and urbanicity are all predictive of one or more outcomes from the NCVS, and are candidates for the underlying factor in the common cause models. Additional analysis shows that these effects persist even after controlling for demographic and design factors. However, the relationships implied by directions of the associations between the underlying factors, coverage propensity and the key outcomes, do not indicate a systematic bias resulting from housing unit coverage in NCVS data that would be of concern.

Exploring Nonresponse and Coverage in a Web Study

Ipek Bilgen, *NORC at the University of Chicago*
Michael J. Stern, *NORC at the University of Chicago*
David Sterrett, *NORC at the University of Chicago*

Researchers are increasingly conducting multi-mode surveys that include a web component either concurrently or subsequently with alternative modes of data collection. Additionally, the

AAPOR 70th Annual Conference

rise in web enabled mobile devices such as smartphones and tablets makes the web surveys an even more attractive mode for surveying the general population. Still, while web surveys have been expected to be a viable data collection mode, the increase in response rates to web surveys has not accelerated as it was expected. Among the households with internet access, the reasons why people respond to surveys via conventional modes (e.g. mail and phone) rather than web is related to both coverage (e.g., lack of sufficient internet speed or proficiency) and nonresponse (e.g., lack of time, interest, or other reason). This paper seeks to examine the reasons for nonresponse to web surveys and disentangle web-specific coverage and nonresponse errors. In order to do so, we have followed up with nonrespondents to an Illinois health survey in which we sent 4,000 sampled Illinois addresses a series of survey invitation mailings including a link to the web survey. In the follow-up study, we first mailed out a self-administrated questionnaire (SAQ) booklet and then administered the questionnaire via telephone interviews (CATI) for 4 weeks approximately three weeks after the SAQ mailings. In both modes, we asked sample members a series of questions regarding their electronic device ownership and mobile internet access, as well as questions regarding whether respondents remember receiving the web invitation letter or e-mail, and their reasons for not responding if they remember such invitations. The key rationale of this paper is to understand the reasons of web nonresponse among individuals with internet access. This will enable us to tailor our data collection strategies to drive more households to the less-expensive web mode.

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

Creating an Electronic Cognitive and Usability Testing Protocol: Decision-Making and Tradeoffs

Libby Snow, *NORC at the University of Chicago*

Rene Bautista, *NORC at the University of Chicago*

One area of survey research that commonly poses challenges is the development of the research protocol for conducting cognitive interviews. Survey methodologists need to weigh many factors, ranging from cognitive theoretical aspects to costs and timing. Cognitive research protocols can be designed under different theoretical approaches, including think-aloud, concurrent probing and retrospective probing, and the interviews themselves can utilize proactive or reactive probing. While survey methodologists have widely adopted a four-stage cognitive model to understand psychological mechanisms around survey answering (Tourangeau, et al., 2000; Willis, 2005), there are fewer practical discussions in the literature on how to design a research protocol, and how such design can impact findings. This paper describes the decision making process involved in designing a cognitive and usability testing protocol for a large scale survey that measures physical and mental health of children in the United States; namely the National Survey of Children's Health (NSCH). The NSCH (which collects data on children both with and without special health care needs) has historically been fielded as a telephone survey, but is looking to transition to self-administered modes (i.e., web and mail). As part of the questionnaire redesign process, NORC conducted cognitive interviews among the general public. NORC developed an electronic system (i.e., Access database) to collect data in real time and guide interviewers through the cognitive and usability testing process. We discuss aspects related to the selection of approaches that work with an electronic

protocol that may be useful for practitioners in the area of health surveys. We also discuss next steps for an electronic protocol.

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

Reducing Skip Errors: Analysis of Visually Grouping Skip Patterns by Question Type

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

Mail surveys continue to be popular among researchers as they can achieve better coverage (Iannacchione, 2011) and response at a lower cost than telephone surveys. However, in mail surveys, respondents have to correctly navigate the questionnaire including all skip patterns on their own. Because of this, researchers need to understand how to design skip patterns to assist respondents and reduce skip errors. Reducing skip errors can help increase response quality and statistical power on questions that are part of a skip pattern. A previous study found that using separate backgrounds for the feeder and follow-up questions to create proper visual grouping decreased skip errors while using indentation to establish visual grouping increased skip errors (Powell, 2014). However, this study only examined skip patterns with behavioral feeder questions, and many of the follow-up questions were open-ended. In this paper, I examine the effect of visual grouping within skip patterns on both behavioral and attitudinal feeder questions, as well as both open and closed ended follow-up questions. The data come from a survey of 3,000 Nebraskans conducted from July through October 2014 (n=1,082; RR1=36.1%). It contained a 2x2 factorial experimental design examining the effectiveness of separate enclosure and indentation for follow-up questions on the reduction of skip errors. Preliminary results show similar overall patterns as the previously mentioned study, but once the type of question was accounted for, the effect of the visual grouping was reduced.

Cross-Cultural Perceptions of Youth: A Comparison Between the United States and Syria

Laura Allen, *University of Nebraska-Lincoln*

This study examines the opinion of the American and Syrian publics regarding their perception of the opportunities of, and respect for, youth. With the uprisings of Middle Eastern youth, including Syria, there appears to be a disparity and disagreement between the younger and older generations. Opportunities for education and employment are of great concern for Syrian youth, causing them to take action through resistance and negotiation with authority (Buckner & Saba, 2010; Theodoropoulou, 2012). While this has also been of concern for American youth, there have been initiatives in the United States to address these issues and demonstrate the priority placed on resolving these issues (Catalano et al., 2004). Because of these types of initiatives and the lack of uprising of American youth, it is expected that older American respondents have a more positive perception of youth than do older Syrian respondents. Using data from the 2011 Gallup World Poll, a multinational probability-based survey, this study will examine the differences in youth perception between Americans and Syrians. Perception of youth will be evaluated based on answers to questions about beliefs regarding opportunities for

AAPOR 70th Annual Conference

children and youth and on respect for children and youth. This will be compared across perceptions of the local economy and household income. Preliminary analysis indicates that U.S. respondents view opportunities to learn and grow for youth more positively with 83.34%, as compared to 76.43% for Syrian respondents. The data also indicate that U.S. respondents view their local economy more positively with 48.12% reporting they believe their local economy is improving compared to 25.03% of Syrian respondents. Understanding the dynamics of public opinion regarding youth could lead to interventions that may lessen generational disparity.

Effects of the 2010 Haiti Earthquake on Haitian Public Opinion

Devin Van't Hof, *University of Nebraska*

Opinions among a population are generally slow to change, but when opinions do change quickly, it can reveal valuable knowledge for researchers and policy makers. One instance that may foster rapid opinion change is in the case of natural disasters. In 2010, Haiti was devastated by a 7.0 magnitude earthquake and a faltering government response. Unfortunately, this happened to a country that already lagged well behind other Western countries in development. According to the United Nations Human Development Index, in 2010 Haiti was ranked 167 out of 187 nations. This combination of events gives researchers a unique opportunity to study how negative experiences may impact a population's trust in others. Previous research has investigated the economic impact of natural disasters on both a short-term and long-term bases (The World Bank, 2014). In addition, the United Nations has described the need for the safety and wellbeing of natural disaster victims. This study examines how public opinion has changed, examining Haitian's attitudes and opinions before and after the 2010 earthquake. Specifically, we concentrate on Haitians' trust in their neighbors and government, and track the optimism of the population. Using data from the 2008 and 2010 Gallup World Polls, probability-based multi-national surveys, this paper examines Haitian opinions before and after the January 12, 2010 earthquake. Preliminary analysis shows that the percentage of Haitians who had confidence in their government dropped from 22.0% before the earthquake to 15.9% after the earthquake. Additionally, in 2008 38.2% of Haitians reported that their standard of living was improving versus 27.4% in 2010. This study helps us better understand the importance of a prompt response from national governments and aid organizations in the immediate aftermath of a natural disaster.

Assessing Mental Health Care Disparity Using Behavioral Risk Factor Surveillance System Data

Jin Liu, *University of South Carolina- Columbia*
Zhihong Fan, *Center for Disease Control*

Modules or questions from the Behavioral Risk Factor Surveillance System (BRFSS) can be used to assess mental health care disparity. Frequent Mental Distress (FMD), defined as having 14 or more mentally unhealthy days during the past 30 days, is one of the 4 items in CDC Health-Related Quality of Life scale. The scale was implemented in all states since 1993. The Patient Health Questionnaire-8 (PHQ8) that is used to screen for current depression (CD) or major depression (MD) was administered in 2010 for some states. Kessler-6 (K6) which is a screener for serious psychological distress (SPD) was administered in 2009 for some states. Receipt of mental health diagnosis and mental health treatment was ascertained simultaneously as part of the respective module in 2009 and 2010. Researchers are concerned whether these modules/questions can equally assess mental health care disparity in demographically diverse populations. The data was analyzed using log linear regression (SAS 9.3). Among adults who experienced FMD, CD or MD in 2010, the following subpopulations had a lower likelihood of

AAPOR 70th Annual Conference

obtaining diagnosis of depression: men, non-Hispanic blacks, and adults 18 to 24 and 65 or older, currently employed, and currently married. Among adults who experienced FMD or SPD in 2009, the individuals with SPD from see the following subpopulation groups had lower likelihood of receiving mental health treatment: men, adults aged 65 or older, non-whites, persons unable to work, and persons without health coverage. Replication of analysis with FMD which is a single-item question actually reached similar conclusions. PHQ8 and K6 are multi-item psychometric scales and burdensome to administer in a comprehensive public health survey among general populations. The likelihood of missing data caused by withdrawal and invalid responses were high. In comparison, FMD appears to be an effective alternative to PHQ-8 and K6 in assessing mental health care disparities.

“Screen and Go” Interviews and Broken Appointments in a National Survey: Trends and Implications

Maura Bardos, *University of Michigan*

Heather Schroeder, *University of Michigan*

Brady T. West, *University of Michigan*

Examining outcomes of interviewer-respondent interactions can give key insights into practices that may increase response rates while minimizing data collection costs. This study examines two possible outcomes of interviewer-respondent interactions in the National Survey of Family Growth (NSFG): “screen and go” interviews and broken appointments. In the NSFG, interviewers conduct a short screening interview to determine if any persons age 15 – 44 reside in a sampled housing unit. For households with more than one eligible person, the CAPI sample management system randomly selects the respondent. A “screen and go” interview occurs when the selected respondent is home during the screener and completes the main interview in the same visit. These types of interviews are ideal as they minimize interviewer follow-ups; however, they occur infrequently and comprise 9 percent of all screening interviews in households with eligible persons. In cases where the selected respondent is either not home or does not agree to complete the main interview, interviewers are encouraged to schedule an appointment to return later, as respondents who make appointments are known to have a higher propensity to respond than those who do not. While over 40 percent of selected respondents make at least one hard appointment, about one in three appointments is broken, requiring additional interviewer effort to complete the interview. We present rates of broken appointments and “screen and go” interviews over time and profile the demographics of respondents who are most likely to exhibit these practices. We also examine the effects of an ongoing incentive experiment on both outcomes. Finally, we compare the relative effort rates required by interviewers to complete an interview in different scenarios (“screen and go,” broken appointments, kept appointment, and no appointment) and present cost estimates of the differences. We conclude with a discussion of implications for future data collection efforts.

Reversed Spiral of Silence: A Case Study of Fang-Han Debate on the Outspokenness of Deviant Opinions on the Internet in China

Qian Liu, *Department of Media and Communication, City University of Hong Kong*

Spiral of Silence refers that people’s outspokenness would be affected by the perceived opinion climate and those who perceived themselves as minorities would keep in silence. In China, the Internet has functioned as a platform for people to speak out rather than remain silent for (1) a sense of perceived safety and collective support provided by computer-mediated communication, (2) a distrust in China’s official and large private profit-making mass media organizations, (3) psychological rebellion against the repressively political and social

AAPOR 70th Annual Conference

environment, and (4) high attitude strength, which has made the spiral of silence a reversed one. People would speak out regardless of the perceived opinion climate, even when they perceived themselves as minorities and hold deviant opinions against the majority view. This study examines how the above factors interacted with each other and affected the outspokenness of deviant opinions of Chinese citizens on the Internet regardless of the media as well as opinion environment, based on the text analysis of participants' posts through different Internet-platforms on Fang-Han debate [Han Han, a famous Chinese writer, racing driver, most influential Chinese opinion leader and public intellectual who was chased by millions of fans, especially among youth and adolescents, was questioned on whether he was faked and all his writings were written by his father and a group of ghost writers, by a few intellectuals (headed by Fang Zhouzi, a well-known Chinese intellectual famous by his efforts in exposing academic frauds) publicly on the Internet, which caused a heated debate lasting several months in 2012 (still on-going now but not so heated) among Chinese opinion leaders, scholars, professors, journalists, intellectuals from different industries, college students, adolescents and so on.].

The Benefits of an Increased Cellphone Allocation to Target Low Socioeconomic (SES) Persons

Jamie Ridenhour, *RTI International*

Marcus Berzofsky, *RTI International*

Bo Lu, *The Ohio State University*

Caroline Blanton, *RTI International*

G. Lance Couzens, *RTI International*

Kimberly Peterson, *RTI International*

Timothy Sahr, *The Ohio State University*

Robert Ashmead, *The Ohio State University*

Amy Ferketich, *The Ohio State University*

Tom Duffy, *RTI International*

Research has shown that the rate of movement to majority cellphone use from partial or full 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 costs for samples drawn from a cellphone frame generally cost more than those selected from a landline frame, the increased rate by which the key subpopulations of interest are found mostly negates the cost per complete difference to obtain the same number of completes. 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 as a guide for similar dual-frame surveys.

Do Late, Reluctant Respondents Give Poor Data Quality?

Herschel Lisette Sanders, *University of Maryland*

Survey participation rates have been decreasing over time. Due to nonresponse bias concerns, efforts have been exerted to get higher response rates. Based on the total survey error

AAPOR 70th Annual Conference

framework, Groves (2004) posited that unintentional increases of other types of survey error may occur when attempts are made to reduce another type of survey error. Using the Home Energy Use Survey 2014, we examine whether by increasing response rates through more recruitment effort, we are unintentionally lowering our overall data quality. Data quality will be measured by the number of items missing. The Home Energy Use Survey 2014 is a multi-stage dual-mode survey. Respondents are first recruited to respond to the web survey. After some time has elapsed, if they haven't responded to the web survey, they will be sent a mail survey. Respondents are also asked a battery of questions relating to their reluctance in giving out their data. Items missing were calculated in the aggregate level for questions asked for all respondents. The relationship of items missing with being late and/ or reluctant will be explored.

Framing the Conservative Case for Same-Sex Marriage: Atypical Republican Issue Advocacy and the Effects of "Conflicting Cue" Value Frames on Same-Sex Marriage Attitudes Within the Republican Party

Ashley A. Koning, *Rutgers University*

As public opinion overall increasingly supports same-sex marriage, the Republican Party continues to expressly oppose it. But some GOP members have been bucking their party's position recently and acting as "atypical issue advocates" for same-sex marriage legalization. These Republican advocates moreover express their support by framing their messages not in the same context as their left-leaning counterparts but rather around the very 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: traditionally conservative values are used to frame support for more progressive issue positions. We know messaging lies at the heart of politics, but we do not know if this particular type employed by Republican atypical issue advocates – what I call "conflicting cue" value frames – is effective, particularly when many Republicans hold strong preexisting beliefs about same-sex marriage, making it potentially difficult to evoke any attitudinal shifts. Existing research on value frames and their effect on public opinion has thus far focused mainly on consensus values or on partisan values supporting established partisan positions. Building on this literature, this paper introduces the concept of "conflicting cue" value frames, in which long-held partisan values are attached to atypical party stances. I analyze through survey experiments the value-laden arguments currently being used by Republican advocates in favor of same-sex marriage and assess their impact on attitudes both within and across partisanship. These frames may have the ability to move opinion within groups where opinion was once thought of as immovable, showing the importance resonant values can play. As the GOP stands at the brink of reassessment and potential "civil war" over social change, this research strives to untangle a new kind of framing in real world politics.

Partisan Flocks: The Influence of Congregation on Vote Choice

Steven Andrew Snell, *Duke Initiative on Survey Methodology, Duke University*

How do congregations shape religious Americans' voting decisions? Political behavior research most often conceives of religious influence in terms of broad categories of religious affiliation, thereby shunting attention away from religious congregations, the local religious communities with which a majority of Americans participate. In contrast, the present work considers how political norms that arise within religious congregations constrain congregants' decision-making. I draw on a nationally representative sample of American congregations in order to demonstrate that congregational partisan norms are prevalent and that there is significant variation in congregational norms across and within religious traditions and denominations. I also employ

AAPOR 70th Annual Conference

data from two original surveys of religious voters to evaluate the extent to which partisan norms at the congregation level influence congregants' vote decisions. I further show that norm constraint is strengthened by two social factors: embeddedness and surveillance. Finally, as a further test of the relationship between congregational norm and congregant political behavior, I turn to two priming experiments, which use question order to increase the salience of religious congregation to voters. I show that those primed to think about their religious affiliation are more likely to report a vote decision consistent with their perception of their congregation's political norm. Nevertheless, this finding holds only for congregants in perceived right-leaning congregations.

The Polarization of Global Climate Change Attitudes: Differential Effects of Comedy, Science and Political Media Content

Hyoyeun Jun, *University of Georgia*

Hanyoung Kim, *University of Georgia*

Michael A. Cacciatore, *University of Georgia*

Work on public opinion of climate change has focused, for instance, on the predictive power of political partisanship (Pew Research Center, 2009), the impact of different labels for the phenomenon (Schuldt, Konrath, & Schwarz, 2011), and more recently, how factors like trust mediate the relationship between media use and attitudinal evaluations (Hmielowski, Feldman, Myers, Leiserowitz, & Maibach, 2014). Nevertheless, our understanding of the opinion dynamics surrounding climate change can be best described as murky. In this study we seek to fill a gap in the literature by exploring the impacts of different types of media consumption on citizen opinions toward climate change. As part of a project at the [removed for blind review], we collected survey responses using a probability sample of registered voters (N = 220) in Oconee County, Georgia. First, our results corroborate previous research that found political ideology to be a major driver of climate change attitudes. Second, we examined the role that media played in predicting such attitudes. Increased attention to national political media was associated with a lower likelihood of reporting that climate change was caused by human activity, while attention to late night comedy shows, including *The Daily Show* with Jon Stewart and *The Colbert Report*, was associated with a greater likelihood of believing that climate changes was human-caused, that it was leading to more frequent and intense natural disasters, and that the U.S. government should take action to halt its effects. Interestingly, attention to science media content failed to predict attitudes across any of our dependent variables. The overall pattern in our data was one of widening opinion gaps between frequent and infrequent users of these media. The implications of these findings, particularly as they relate to future public opinion patterns and to the issue of attitude polarization for this topic are discussed.

The Impact of Income-Partisan Dealignment on Economic Voting

Ju Yeon Park, *New York University*

The impact of the economy on vote choices has been studied in two ways: voting on economic conditions and voting on economic policies. When both voting behaviors were tested vis-à-vis, economic conditions were found to have a stronger effect. Given that some economic policies often benefits one class at the cost of the others, this paper investigates whether economic voting weakens as the alignment between income groups and partisans strengthens. Based on a multi-level analysis using survey data from the Comparative Study of Electoral Systems covering 40 countries and 97 elections with 120,911 individual responses in 1996-2011, I find that affluent voters vote on economic conditions when income groups and partisans are perfectly dealigned but do not show economic voting as they become aligned. On the other

AAPOR 70th Annual Conference

hand, poor voters react only to the unemployment rate and, interestingly, punish the incumbent for an increasing unemployment rate more when the partisan alignment is strong.

Revisiting the Measurement of Partisanship - A Latent Trait Approach

Alexa Bankert, *Stony Brook University*

Leonie Huddy, *Stony Brook University*

In the study of voting no concept has received as much attention as partisanship. Since the notion of party identification was introduced about sixty years ago, scholars have continuously debated the concept and its measurement. Although the original survey items are still widely used, several scholars have emphasized that these do not link up well with the underlying theory. In addition to this theoretical disconnect, the traditional partisanship measurement also lacks scale qualities that we find necessary given the concept's central position in political science. In this paper, we suggest the measurement of partisanship as a social identity. We examine the validity and reliability of a multi-item index of partisanship on the basis of German, Dutch, Swedish and U.S. survey data. Utilizing an Item Response Theory framework, our findings show that such an index has better measurement properties and constitutes a stronger predictor of political attitudes and political participation than the traditional measure. We also identify a set of items to use in a shorter version of the index for measuring partisanship.

Methodological Considerations in the Use of Name Generators and Interpreters

David E. Eagle, *Duke University Department of Sociology*

Rae Jean Proeschold Bell, *Duke University Global Health Institute*

With data from the Clergy Health Initiative (CHI) Longitudinal Survey, we look for interviewer effects, differences between web and telephone delivery, and panel conditioning bias in an "important matters" name generator and interpreter, replicated from the U.S. General Social Survey. The CHI survey data on approximately 1700 respondents per wave was collected in 2008, 2010, and 2012 on all currently serving United Methodist Clergy in North Carolina with response rates of 95%, 87%, and 81%. We find evidence of phone interviewers systematically influencing the number of confidants named, we observe that respondents assigned to the web survey reported a larger number of confidants, and we uncover strong support for panel conditioning. We discuss the possible mechanisms behind these observations and conclude with a brief discussion of the implications of our findings for similar studies. This studies raises important considerations for the design of social network collection instruments and how the visual design of web surveys plays an important role in shaping responses to name generator questions.

Is Mandatory Drug-Testing for Welfare Recipients Grounded in Evidence or Public Sentiment?

Lillian Reed Walsh, *East Tennessee State University/Applied Sociological Research Laboratory*

Nikki M. Bare, *ETSU*

There are several legislative measures pending across this country that would mandate drug testing for welfare recipients. Supporters of these policies say that they are ensuring that tax dollars are allocated to people who are using them appropriately. Opponents worry that these policies are not actually being used to reduce drug use but rather as a mechanism by which they can strip the poor of their government assistance. The underlying message behind these

AAPOR 70th Annual Conference

policies is that welfare recipients are lesser individuals who are living off government assistance with no desire to get a job and get off government aid. However, is this assumption driven by data or something else? This paper will explore existing data to determine validity, relevance, cost, consequence, and need for such a controversial procedure. Furthermore, it will look at the impact of public opinion in shaping the debate of these policies by examining the individuals calling for these tests and upon what evidence they base their claims. Given that these policies would impact some of the most vulnerable in our population it is important understand the motivations behind these policy decisions. The goal of this project is to add to a larger dialog about the implications of drug testing on vulnerable populations or those populations without the ability to refuse the drug tests. The discussion will center around evidence-based practices that actually reduce drug use versus policies that may have very vocal support among some but are not grounded in scientific evidence.

Offline Data Collection in Sub-Saharan Africa Using SMS Surveys: Lessons Learned

Carsten Broich, *Sample Solutions*

Daniel Boonman, *Sample Solutions*

According to the International Telecommunication Union (ITU) cell phone penetration in large parts of Africa has reached levels beyond 100% penetration. A question however remains which is how these respondents can be included in public opinion polling since in many cases the population uses cell phones consisting of basic functionality with call and text messaging options. The large penetration of smartphone or so-called 'feature phones' as they are called in many Asian regions, is yet to come. Subsequently an issue for data collection remains in the fact that large parts of the population do not have access to the internet. In many urban areas the usage of internet hubs such as internet cafe is in many locations the predominant way to access and exchange information. This also has a large influence on survey methods that can be established in these areas. Nevertheless as SMS services are very cost-effective in Sub-Saharan Africa it could serve as a vital tool for ad-hoc omnibus study on a national level. Within this study the setup of a national survey using cell phone respondent as sampling frame is outlined and hereafter the fieldwork progress and its results are discussed. Implications such as non-response bias, incentives and reminder notifications are also discussed. Furthermore improvements techniques to overcome some of these issues are being outlined. Lastly a respondent check is conducted with which the ownership change rate in Sub-Saharan Africa is analyzed. Some of the aspects includes the sampling frame but also aspects like financial viability.

Can Labeling Participants in a Survey Cover Letter Affect Response Rates?

Jessica Jordan Sykes, *East Tennessee State University*

Kelly N. Foster, *East Tennessee State University*

Labeling people can have a serious effect on how they see themselves and how they respond to us as researchers. Some labels have a negative connotation, some have a positive one. In fall 2014, the Applied Social Research Laboratory (ASRL) at East Tennessee State University is sending a survey to adults over 50 to gather their opinions about a website designed to connect senior citizens with services they may need or want. Cognitive interviews and background research suggests that older adults, particularly those who are still very active, do not like being called "senior citizens" and would prefer for that not to be part of their labeling. This experiment aims to test how the act of labeling participants in a cover letter for a survey could potentially affect the response rates. We will use two different cover letters with a paper survey being sent

AAPOR 70th Annual Conference

via standard mail to individuals age 50 and above in a rural Appalachian area regarding information they would like to see in a senior services website. One cover letter will explicitly use the term “senior citizen” while the other will say “adults over age 50”. It is our hypothesis that the term “senior citizen” evokes images that the individuals do not associate with themselves, thereby increasing the likelihood that they will be turned off by the survey and not complete it. We anticipate a higher response rate in the population that receives the “adults over 50” wording versus “senior citizen”. It is our hope that this project will serve to inform ways in which we can engage the aging population by utilizing language that is appropriate and that they feel encompasses their identity.

Linking Mindfulness to Social Desirability Bias: Do Mindful Respondents Tend to Give Socially Desirable Answers?

Ji Qi, *University of Michigan Institute for Social Research*

Ting Yan, *Westat*

Florian Keusch, *University of Mannheim*

James Lepkowski, *University of Michigan Institute for Social Research*

David L. Vannette, *Stanford University*

Social desirability bias (SDB) is considered as an important source of measurement error for surveys asking sensitive questions. Previous literature in survey methodology as well as social psychology has provided substantial and valuable insight on causes of SDB and methods to reduce SDB. However, little attention has been given to an individual person's level of mindfulness—the tendency to stay focused during a specific task or in general— as a potential factor influencing the occurrence of SDB. This paper investigated the associations between respondent's mindfulness and their answers to two sensitive questions measuring attitudes towards gays and lesbians and illegal immigrants. In addition, we examined the potential moderating effect of question order (questions are either asked in an increasing order of sensitivity or a decreasing order of sensitivity) on the relationship between mindfulness and socially desirable responding. Bivariate analysis revealed a significant positive effect of mindfulness on socially desirable responses, indicating that mindful people are more likely to provide socially desirable answers than those less mindful. However, the positive effect turns statistically non-significant after controlling for gender, age, race and educational attainment. We didn't find a moderating effect of question order on the relationship between mindfulness and socially desirable responding. This paper provided the first empirical evidence linking mindfulness to social desirability bias. We will discuss the implications of our results for survey researchers and talk about next steps to carry this research further.

How Does Your Family Use Mobile Devices? : A Case Study on the Proxy Response

Vera Kurmlavage, *The Nielsen Company*

Kumar Rao, *The Nielsen Company*

Anh Thu Burks, *The Nielsen Company*

Jennifer Haskell, *The Nielsen Company*

Proxy reports in a survey are responses provided by a respondent about another member of the sampled unit or household. While a common practice in Government surveys (ex: Current Population Survey and National Health Interview Survey), proxies are also used in health surveys and epidemiological studies to substitute for the absence of the chosen person at the time of the interview or inability of the chosen person to answer. However, there exists little or

AAPOR 70th Annual Conference

no evidence in the literature when it comes to a proxy response's viability and use in surveys conducted by research companies on topics that are more germane to consumer activities and behaviors (such as media consumption). In this study, we analyze the convergence and divergence of self and proxy reports across a variety of behavioral questions; specifically, how members of a home use mobile devices (ex: Tablets and smartphones). The data come from self-administered (proxy) and supplemented interviewer-administered (self) surveys conducted on members of a probability-based panel in the U.S. The research objective was to assess the accuracy of proxy responses and in the effort, understand the social and cognitive processes that underlie the divergence of self and proxy reports. Specifically, exploring how people store, integrate, and recall information about the self and the behaviors of others. Findings revealed that self and proxy reporters process and respond to questions differently. Divergence between the two response types was a function of "data specificity" i.e. questions that ask for specific details about the behavior of another person. Interestingly, certain characteristics of the proxy respondent (age, family relationship, etc.) also played a moderating role in the convergence of proxy and self-responses. The findings from this study not only extend the body of knowledge on self-and-proxy responses, but also facilitate in providing practical guidelines for questionnaire construction and respondent selection.

How to Maintain an Internal Do Not Contact (DNC) List? Findings the Best Practices for Companies and Organizations

Yelena Pens, *The Nielsen Company*

Robert DeHaan, *The Nielsen Company*

Colin King, *The Nielsen Company*

In recent years there has been an increase in legal concerns regarding do not contact (DNC) lists, opt out lists, and many alleged violations of the Telephone Consumer Protection Act (TCPA) (e.g., calls to wireless numbers without consent, using auto-dialers to call wireless numbers and pre-recorded message blasts). Nielsen, a leading global information and measurement company, has an obligation to honor the request of the respondent and maintain an opt-out list which includes do not call, mail, visit, and/or email. However, are there industry-wide best practices for maintaining internal DNC lists? With no industry approved practices, research companies and other organizations have to develop company specific rules for adding and removing people from their list. Currently there are multiple services representing Nielsen, and each service has separate business rules for placing contacted households on DNC list. The lists range from a few thousand households to several million. Nielsen has expressed legal concern over a higher risk having different business and interviewer calling rules between services that represent the same company. The objective was to determine the best practices for business rules that will govern "do not contact" procedures for all services as well as conduct an impact analysis on the implications to key performance indicators. The research plan included comparing business rules such as how long the contact should remain on the list, and conducting an impact analysis on response rates and other quality metrics. What is the balance between addressing legal concerns and maintaining response rates? We will present the results from the impact analysis. Also, we will show the costs and benefits of the selected approach, as well as, present the best practices for creating an integrated service.

Money Matters: How Financially Literate Are Our U.S. Teens?

David C. Miller, *American Institutes for Research (AIR)*

Teresa Kroeger, *American Institutes for Research (AIR)*

Over the past decade—especially since the U.S. subprime mortgage crisis ushered in the Great Recession—there has been increasing concern about people’s level of financial literacy. Early in 2008, an Executive Order was signed creating, for the first time, a President’s Advisory Council on Financial Literacy to improve financial literacy among all Americans. The Council defined financial literacy as “the ability to use knowledge and skills to manage financial resources effectively for a lifetime of financial well-being.” There is particular concern that young people—who may bear more financial risks in adulthood due to increasing college costs and debt from student loans, uncertain economic and job prospects, a decrease in welfare and pension/retirement benefits, and increased life expectancy—are lacking financial literacy (OECD 2014). It is believed that an increase in financial literacy skills can improve financial decision making, and thus positively impact not only individuals and households but also economic and financial stability more generally. In response to this, in 2012 a large-scale international study to assess the financial literacy of young people was launched for the first time. In 2014, the OECD released key international findings from the PISA financial-literacy assessment of 15-year-olds, which the United States participated in along with 17 other countries and economies. This analysis will look more in-depth at the U.S. data, including U.S. students’ performance on the financial-literacy assessment and their responses to survey questions about their frequency of discussing money matters, savings and spending habits, having a bank account, having a pre-paid debit card, etc. Bivariate analyses will look at differences along these variables by demographic characteristics and multivariate analyses will be performed controlling for demographic characteristics. Thus, this poster will not only provide general survey results, but also identify unique predictors of teens’ attitudes, practices, and knowledge about money matters.

Comparing Methods for Correcting Nonresponse Bias in a School Climate Survey

Elisabeth Davis, *American Institutes for Research*

Daniel Harwell, *American Institutes for Research*

Samantha Neiman, *American Institutes for Research*

Sandy Eyster, *American Institutes for Research*

School climate surveys are becoming an increasingly popular tool to measure the environmental factors within schools that can promote or hinder student learning – such as perceptions of safety and academic support. These surveys are typically administered in-school and response rates are, on average, relatively high. Despite these response rates, it is possible that the students who either do not give consent or are not in-school to complete the survey may systematically differ from the students who do so. This research builds on previous work about response propensity to the Conditions for Learning (CFL, a school climate survey developed by the American Institutes for Research) that found that students who were less engaged in their school community, which is a key measure for the CFL, were less likely to complete the survey. The current study utilizes data collected from three administrations of the CFL during the 2013-14 school year to examine the primary determinants, including achievement and attendance data, and the impact of nonresponse using a Chi-Squared Automated Interaction Detection (CHAID) analysis. Additionally, we compare two methods of correcting for nonresponse bias in the data: 1) imputing data from nonrespondents utilizing their responses from another survey administration and 2) using nonresponse-adjusted weights. We compare scale scores for the

AAPOR 70th Annual Conference

four CFL composites and the associated standard errors to determine the impact for each of these methods.

County Level Targeting in the 2014 Maryland Healthier Communities Survey (MHCS)

Marilyn I. Wilkinson, *Abt SRBI, Inc.*

Rachel Martonik, *Abt SRBI, Inc.*

Robert Fiedler, *Maryland Department of Health and Mental Hygiene*

Courtney Kennedy, *Abt SRBI, Inc.*

The 2014 Maryland Healthier Communities Survey (MHCS) is a dual frame RDD telephone survey being conducted in 2014 and 2016. MHCS is using frame-specific targeting strategies to obtain county-level estimates for smoking and other chronic diseases. ZIP code is being used to target landlines, and with the exception of two counties where there are no rate centers within the county borders, rate centers are being used to target cell phones. This poster will compare and assess the landline and cell frame county-level targeting results obtained by the 2014 MHCS.

Let's Talk About Sex: Achieving High Response Rates When Asking Teens About Sexual Behaviors

Meredith Kelsey, *Abt Associates*

René E. Nutter, *Decision Information Resources, Inc.*

Unique challenges exist when conducting surveys with young adults, particularly those with a longitudinal design and focused on sensitive topics. There are concerns about privacy, truthfulness, and retention (Pierce and Hartford). Further, young adult populations are often hard to reach and maintain contact (Cotter et al). The proposed poster will reveal the successful data collection efforts of the Teen Health Empowerment Study, an evaluation of teen pregnancy prevention interventions, which to date, has achieved response rates at or beyond 85% at baseline and follow-up. All data is collected utilizing a relatively new web-based, audio computer-assisted self-interview (ACASI) technology (chosen to control for varying reading levels and privacy). Data collection has occurred at baseline, short- and long-term follow-up periods in the nine sites. The short-term follow-up surveys occur six, nine, and 12 months after baseline, depending on the model. The long-term follow-up surveys are administered 12, 18, and 24 months after the baseline survey through various venues—groups at schools, individually (assisted or unassisted), and in small groups at local eateries or community centers. Although this evaluation has a strong and rigorous design and includes a methodology to address the major concerns when collecting data on sensitive issues, many challenges have arisen. The proposed poster will focus on lessons learned during the data collection activities thus far. Specifically, we will provide details on: · the processes for obtaining parental consent; · branding the study (and including youth-friendly materials, t-shirts, and communications); · creatively tracking potential respondents; · locating teenagers with virtually no “electronic footprint”; and · technology issues with web-based data collection. Each of these things has contributed to the success of this evaluation to date, and the lessons learned from this poster will assist evaluation teams when strategizing effective methods for collecting data on sensitive topics from hard to reach samples.

Geographical Predictors of GPS-Based Survey Response Rates

Dara Seidl, *Abt SRBI*

Timothy Michalowski, *Abt SRBI*

Geographic Information Systems (GIS) and geostatistics have been shown to have utility in assessing the predictors of response rates in survey research. When controlling for demographic variables, geography can impact response rates when the same study is conducted across regions. As the burden increases for survey respondents in terms of the tasks they must perform, there is an expected drop-off rate in full study compliance. In GPS-based surveys where GPS data loggers are deployed to recruited households, the households are expected to both record their locational data and return the units for full compliance. Earlier research has demonstrated that the number of household members is a significant negative predictor of compliance in surveys with GPS devices when a single incentive is applied per household. The willingness to both return the GPS units and to complete the study with data are indicative of full compliance with the survey. This study examines three levels of compliance in GPS-based surveys: (1) mailback of GPS devices, (2) mailback of GPS devices with data, and (3) mailback of GPS devices with data recorded on every day of the study period. This research leverages spatial statistics and regression metrics to compare how household compliance varies across regions when controlling for mean household age and number of persons in the household. A second component tests how well compliance can be predicted with geostatistical interpolation tools in a single study area based on the response rates of a pilot in the same region. Predictions are validated against the geographical compliance patterns of the main surveys. Successful prediction of GPS compliance by geography indicates potential utility in adaptive design measures.

A Novel Approach to Coding Qualitative Survey Responses in a Health Survey: Harnessing the Power of Natural Language Processing (NLP)

Andrew Hurwitz, *Mathematica Policy Research*

Jeremy Biggs, *Mathematica Policy Research*

Kate Dovgala, *Mathematica Policy Research*

Martha Kovac, *Mathematica Policy Research*

Natural Language Processing (NLP) is a powerful tool that integrates principles from artificial intelligence and linguistics to generate meaningful inference from human language. NLP relies on programmed algorithms to perform numerous tasks (e.g. machine translation, automatic speech recognition, semantic processing), but most relevant to survey research is NLP's ability to analyze qualitative data and code this data to a pre-determined coding scheme. This application of NLP is referred to as text classification and has been used by researchers in other fields. One application has been to code radiology reports concerning the potential diagnosis of pneumonia in infants (Mendonca, Haas, Shagina, Larson, & Friedman, 2005). The researchers utilized qualitative data provided in these reports and applied a NLP algorithm to determine whether or not the infant had pneumonia. Other researchers utilized NLP to analyze qualitative information from pathology reports to determine minority women's eligibility for participation in a cancer study and cancer prognosis (Xu, Anderson, Grann, & Friedman, 2004). Psychological research has also benefited from the use of NLP. NLP was utilized by Dale and Spivey (2006) to analyze syntactic coordination between children and caregivers. Despite these varied applications of NLP, the literature reveals no attempt by survey researchers to apply NLP to open-ended text response coding. Utilizing data from a health survey of young adult women, we examine the viability of using NLP to code survey responses as opposed to manual human coding. We present a straightforward bag-of-words model, and train this model using a variety

AAPOR 70th Annual Conference

of supervised learning algorithms (e.g. logistic regression, multinomial naïve Bayes, linear support vector machines). Results are reported for sensitivity, specificity, and positive/negative predictive value of the trained models. Finally, we discuss cost considerations and present a cost benefit analysis for survey researchers considering NLP for use in coding open-ended responses in survey instruments.

The Results Are Only as Good as the Sample: Assessing the Accuracy of Three National Physician Sampling Frames

Rachel Kogan, *Mathematica Policy Research*

Catherine DesRoches, *Mathematica Policy Research*

Kirsten Barrett, *Mathematica Policy Research*

Bonnie Harvey, *Mathematica Policy Research*

James Reschovsky, *Mathematica Policy Research*

Bruce Landon, *Harvard Medical School*

Steve Shortell, *University of California, Berkeley*

Lawrence Casalino, *Cornell University Medical School*

Eugene Rich, *Mathematica Policy Research*

The shortcomings of physician sampling frames are well known. However, little is known about how specific sampling frames differ from each other in accuracy and coverage. We compare the National Provider and Plan Enumeration System (NPES), the American Medical Association Masterfile and the SK&A physician file. We sampled 3,000 from the NPES (500 in each of 6 specialties). We conducted two- and three-way comparisons across data files on address fields and specialty to determine the extent to which they matched. In addition, we randomly selected 1,200 physicians (200 in each specialty) for telephone verification. 1,655 physicians (55%) were found in all three data files. Rates of “missingness” were highest in the SK&A data file when compared to the NPES and varied by specialty (50% in radiology vs. 28% in cardiology). Two-way comparisons found NPES and SK&A had the highest rates of matching contact information, while the AMA Masterfile had low rates compared with the NPES and SK&A. We were able to confirm 65% of physicians’ contact information by phone. The NPES and SK&A had similar rates of correct information in phone verification (72%-94% and 79%-91% respectively across specialty), while the AMA Masterfile had significantly lower rates of correct contact information across all specialties (30% -55% across specialties). These findings indicate none of the data files in this study are perfect: the fact that we were unable to reach one-third of our telephone verification sample is troubling. However, the study offers some encouragement for researchers conducting physician surveys. The NPES appears to provide accurate, up-to-date contact information for physicians billing public and provider insurers. Use of this free data file would allow researchers to reallocate financial resources from purchasing sample to other aspects of survey administration that could positively affect response rates and improve the overall quality of physician survey data.

Advancing Towards a Set of Best Practices in Managing Large Scale Qualitative Research Projects

Martha Stapleton, *Westat*

Darby Steiger, *Westat*

Large scale qualitative research projects require a special set of procedures to implement and manage, yet the literature is sparse on theories and best practices for doing so (Beatty and Willis, 2007; Tucker, 1997; Willis, 2005). Sha and Childs (2014) and Sha, Kenward, Feldman

AAPOR 70th Annual Conference

and Heimerl (2012) document a project management approach for a large scale study, which includes innovative techniques such as mailing advance letters to potential respondents; using web-accessible calendars to track interviewer availability for scheduling of respondents; enforcing consistent file naming conventions; and conducting a root cause analysis to improve participation after the first round of interviews. Many of their recommendations can be implemented on qualitative studies of any size, but more studies need to be documented in order to work towards a set of best practices for execution of large scale qualitative research. This paper will present an additional perspective on large scale qualitative study management, based on our experiences conducting three rounds of cognitive testing of the American Community Survey with 420 respondents meeting more than 50 unique recruitment criteria. We developed 27 different protocols in two languages; intensively trained more than two dozen interviewers; interviewed households and group quarters in 8 cities plus Puerto Rico; and trained and supervised multiple teams of coders and analysts across two organizations. We'll discuss in detail the processes we implemented for each of these aspects of the project, along with lessons learned and suggestions that contribute to the formation of best practices.

Mindfulness in the Survey Context: Who are Mindful and Who are Mindless?

Ting Yan, *Westat*

Florian Keusch, *University of Mannheim, Germany*

James Lepkowski, *Institute for Social Research, University of Michigan*

There exists a large body of research on mindfulness and mindlessness in various contexts, such as health, education, business, social issues, and others. The growing literature on mindfulness demonstrates the benefits of being mindful over being mindless in those contexts. Vannette and Krosnick (2014) made the first attempt to link the literature on mindfulness with the literature on survey methodology. However, their attempt remains conceptual and speculative in nature. There is no empirical research yet studying mindfulness and mindlessness in the survey context. This paper presents the first empirical effort looking into the issue of mindfulness in the survey context. In a telephone survey, we measured respondents' mindfulness at the moment of taking our survey. We also measured respondents' dispositional tendency to be mindful in their daily lives. In this paper, we propose to profile respondents based on these two measures of mindfulness. Specifically, we will characterize respondents who showed different levels of mindfulness when taking the survey and will identify characteristics of respondents associated with different levels of mindfulness in their daily lives. We will then evaluate which measure of mindfulness is more important in the survey context by examining response behaviors at different levels of mindfulness. Our results will have significant implications and will advance literature on both mindfulness and survey methodology.

Enough is Enough: The Optimal Number of Contacts for a Multi-Mode Survey

Stephanie A. Beauvais Dennig, *Westat*

Jocelyn Newsome, *Westat*

Kerry Levin, *Westat*

Brenda Schafer, *Internal Revenue Service*

Jose Colon de la Matta, *Internal Revenue Service*

Past studies have shown that the number and form of contacts during survey administration can significantly influence response rates. The classic Dillman's Tailored Design Method (TDM) (Dillman, Smyth, and Christian 2014) approach advocates contacting respondents 4 or 5 times, where each successive contact is different from the preceding contact. It has been empirically demonstrated that each additional contact will result in an increase in the overall response rate

AAPOR 70th Annual Conference

(Hassol et al 2003, Rookey et al. 2012). When plotted as a curve against level of effort or cost, response rates move incrementally towards an asymptote or a plateau. However, in actual survey practice, we rarely observe a plateauing of the response rate. For many reasons, including budgetary and time constraints, more than 5 contacts is typically not an option in “real world” survey practice. As a result, there is minimal evidence in the literature concerning the optimum number of times a respondent to a survey should be contacted in order to increase response rates and still be cost efficient. In other words, when is enough, enough? In an earlier paper, we examined the impact of a sixth contact in the IRS Individual Taxpayer Burden (ITB) Survey, which annually surveys about 20,000 taxpayers across the United States. We found that the sixth contact significantly increased response rates (Levin et al. 2013). In this paper, we will look at the results of an experiment testing the effectiveness of a seventh contact. The experiment examined the effectiveness of a letter prompting respondents to complete the paper survey, a letter pushing the web survey, and no contact at all. We will look at response rate analysis for each condition as well as the cost per additional complete.

Where Does the Platform Matter: The Impact of Geographic Clustering in Device Ownership and Internet Access in Web Surveys

Ilana Ventura, *NORC at the University of Chicago*

Ned English, *NORC at the University of Chicago*

Ipek Bilgen, *NORC at the University of Chicago*

Michael Stern, *NORC at the University of Chicago*

Due to its relatively low cost and flexibility, the internet has become an increasingly attractive platform for survey practitioners. Moreover, the increasing ubiquity of high-speed internet has quelled past concerns about undercoverage of certain populations. Internet access for lower SES populations is known to be disproportionately mobile and thus dependent on smartphone or tablet-friendly instruments for success. As the dependence on mobile devices for internet access correlated with SES status, we would expect a high degree of clustering in web responses by mobile as opposed to traditional desktop computing. At question is how geographic clustering or autocorrelation in web platform and access, specifically traditional computer plus broadband vs. smartphone or tablet with mobile connections, can influence response rate, and substantive data in a web study. If certain parts of the United States tended to respond using given platform correlated with different response rates, degrees of breakoff, or substantive data, there exists a risk of spatial bias. We analyze paradata and substantive responses from a web survey of health in Illinois to understand the geographic nature of the interaction between internet platform, operating system, screen-size, and response. Specifically, use geographic information systems (GIS), local indicators of spatial autocorrelation (LISA), and spatial-lag modeling to visualize and quantify any clustering and its potential impact on survey data and response. While previous research has shown no correlation between broad band access and response rates (Fiorio et al. 2013), the current poster provides an extension by examining the impact of device ownership, mobile internet access, and internet use frequency. Our research is valuable to designers of web surveys who depend on a combination of mobile and traditional computer users.

AAPOR 70th Annual Conference

An App to Facilitate Real Time Reporting of Interviewer Activities: Revelations and Outcomes

Catherine C. Haggerty, *NORC University of Chicago*

Kymn Kochanek, *NORC University of Chicago*

Micah R Sjoblom, *NORC University of Chicago*

How do interviewers spend their time? How can we maximize interviewers' time spent engaged in "productive" activities? Most survey projects ask interviewers to report time spent engaged in broad activities such as training, interviewing and locating. Interviewer time is typically reported once a week in an electronic timecard that requires the interviewer to report gross rather than granular details about their activities. Survey staff examine time reported in combination with call record entries and analysis of outcome codes which occasionally raises red flags unnecessarily, or more often, fails to reveal important details about goings on in the field. We sometimes observe what appears to be little or no work being accomplished but when quizzed interviewers explain in rich detail exhaustive efforts to accomplish their tasks. NORC has recently developed an App that allows interviewers to report their activities in real time and in richer detail. We will report how interviewers are spending their time. We will describe our examination of the processes and protocols interviewers follow and our attempt to streamline administrative type tasks so that interviewers have more time to engage in their core work. We will also share insight gained into the nuanced differences between less and more productive interviewers and the subsequent coaching and mentoring activities developed and executed. Finally, we will report how the real time delivery of the activities interviewers are engaged in has enabled early responsive adjustments so that interviewers are able to work more efficiently and productively.

The Immigration Issue in Mexican Public Opinion: How Does They See Us? How Do We See Them?

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

Jose Alberto Vera mendoza, *Parametria S.A. de C.V.*

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

Mexico is one of the main reserve army of labour in the United States. However, according to new Pew Research Center estimates from the Great Recession of 2007 to 2009, the number of unauthorized Immigrants living in the United States has stabilized and shows no sign of rising. For decades, the immigration issue has been one of the most critical points of the US-Mexico relationship. The Obama administration has not been immune to this problem. The event of unaccompanied migrant children from Central America in the United States has further fueled this discussion. Several studies have shown contrasting views on the issue, even among Hispanics (Pew Research Center, 2014). Also, a recent Gallup report notes that Hispanics remained more likely to name Immigration as the country's top one problems. In contrast, the pollster Parametría shows that for Mexicans the immigration issue has not relevance in the relation to its northern neighbor. The same polling firm released data showing that Mexicans are also divided on the issue of unaccompanied migrant children from Central America. These contrasting views are the focus of this research. Using data from the Pew Research Center and Parametría, this research aims to find significant differences between U.S. and Mexican public opinion on immigration. Recall that Mexico shares a border to the south with Guatemala and Belize. To that extent, this work contrasts the Mexican opinion about its countrymen's migration to America, and the Central Americans' migration to Mexico. This paper serves as a useful tool in the discussion of migration issue in both countries. Identify the points of contrast between

AAPOR 70th Annual Conference

these two audiences, equally Mexicans but with different residence, permits us to identify some features of the discursive shift that requires migration.

Innovations in Registration Based Sampling for Public Opinion Research

Jonathan Robinson, *Catalist, LLC*

As response rates for political surveys fall across the board and non-response and a slew of other biases threaten the reliability and validity of public opinion surveys, it appears that traditional RDD survey methodology is growing inadequate for dealing with these kinds of issues in data collection. In this paper, we show that registration-based sampling (RBS) frames, originating from large-scale databases of registered voters, present the best method for dealing with these biases going forward. Drawing from the national voter file at Catalist, which has been maintained since 2006 and used in a variety of scholarly research (Hersh & Ansolabehere 2012, Rogers & Aida 2014), we show how using tried and true methods from the sampling literature (variants of stratified sampling, including PPS sampling) in concert with large-scale, person-level demographic data better deals with non-response bias and results in samples that more accurately reflect the true distribution of voters on Election Day (something voter file data is better equipped to judge than traditional methods). We use methods that go beyond most of the uses of RBS sampling in academic (Mann et al 2014) and practitioner settings in terms of statistical and computational rigor. Through a set of simulation studies that take advantage of a large database of survey dispositions collected at Catalist, we show evidence for the superiority of these kinds of techniques over the status quo in survey research today. We also touch on how registration-based sampling can enhance the survey weighting process by linking information from surveys not sourced from the voter file through record-linkage algorithms, and provide a set of baseline data to improve weighting.

Forecasting in Low Information Elections: Evaluating an Election Forecasting Model which Aggregates Across Countries to Other Approaches

Julia Clark, *Ipsos Public Affairs*

Clifford Young, *Ipsos Public Affairs*

Neale El-Dash, *Sleek Data*

Forecasting elections is the “bread and butter” of pollsters and political scientists alike. Models range from single polls to the averaging of many polls to the use of “fundamentals” models which predict future elections based on past ones using political and economic factors. More recently, forecasting models have started to combine poll averaging with “fundamentals models”. The use of such combinatorial approaches came into their own during the 2012 Presidential elections. So to what extent can the U.S. experience (with dozens of pollsters and hundreds of polls) be applied to other electoral contexts? The U.S. is a high information electoral environment with a large N size to work with. By contrast, in most other countries, especially emerging markets, very few polls are conducted during the electoral cycle—typically no more than a half dozen polling firms conducting around 40 polls. Additionally, many emerging markets do not have a large sample of democratic elections, as re-democratization for many is no older than a few decades. This is a weakness which frustrates the robust forecasting of elections in such low information contexts. One possible solution with the N size issue is to aggregate across elections and countries. This reduces the problem of N size but presents a new problem: can elections across countries be compared? Our paper has four objectives. First, we will review the literature on electoral forecasting. Second, we will present and examine a forecasting model which aggregates 178 elections across 35 countries. Third, we will compare our forecasting model with other approaches. In particular, we will assess model performance

AAPOR 70th Annual Conference

across 20 different elections in the last 5 years. For this exercise, we will analyze only elections in presidential systems. Lastly, we will discuss different approaches to combining our model to others with the objective of making predictions more robust.

Building a Robust Methodology to Measure Media Ratings Through Mobile Phones

Max Richman, *GeoPoll*

King Beach, *GeoPoll*

Jamal Boubakri, *GeoPoll*

Roxana Elliott, *GeoPoll*

In the US, methods for measuring audience size and TV ratings have been honed over years, and reliable measures of TV and radio consumption are released on a daily basis. However, emerging media markets such as those in Kenya, Uganda, Tanzania, and Ghana struggle with measuring media consumption, meaning broadcasters and brands have had to rely on, at best, months-old data to make programming and advertising decisions. Measuring viewing or listening habits in such a diverse continent as Africa can be difficult, especially when trying to adapt one method for multiple countries. Methodologies that have been successful elsewhere are difficult to implement in regions as large and competitive as those in Africa, and media, especially TV, is often consumed in large groups rather than at the household level. Researchers have also grappled with how to develop a methodology which can be effectively scaled to several markets throughout Africa. This paper presents a new method being used to produce daily media measurement data in Africa: the mobile phone. Mobile phones have greater penetration than landlines and the internet throughout Africa, and are increasingly being used for market research and educational purposes. GeoPoll, a private mobile surveying company, has developed a panel-based survey which measures TV and radio ratings each day by sending users a multi-part text message survey. This paper details how users are selected to take part in the panel, the universe estimate calculation, overcoming difficulties with question wording and respondent shirking, and how this method can easily scale to countries throughout emerging markets. Findings are applicable to both media measurement and general mobile survey research in emerging markets, and include important information on respondent fatigue, sample size calculation, and ratings calculations.

Testing the Effect of an Email Reminder to Reduce Breakoff Rates in the American Community Survey's Internet Data Collection Mode

Rachel Horwitz, *U.S. Census Bureau*

Mary Frances Zelenak, *U.S. Census Bureau*

The Census Bureau's American Community Survey (ACS) added an Internet data collection mode as part of its sequential design in the January 2013 data-collection cycle. Prior to introducing this new mode, the Internet instrument went through two rounds of testing in 2011, which focused on notification strategies, the design of the instrument, and data quality compared to the paper questionnaire. Those tests found higher item nonresponse rates for items later in the survey, as compared to the paper questionnaire, which were related to Internet breakoffs earlier in the survey. Based on these findings, the ACS proposed collecting email addresses from respondents and sending a reminder email to those that had started, but not submitted the survey. This paper discusses the effect of sending a reminder email to breakoff cases. Breakoff cases received one of three emails, which varied in subject line, email content, and the format of the link to access the survey. Specifically, we test whether including the name of the survey in the subject line and body of the email and using a full URL compared to a 'Click

AAPOR 70th Annual Conference

Here' link to access the survey has an effect on email open rates, click-through rates, the return rate of breakoff cases, and item nonresponse rates.

Effects of Acculturation on Question Comprehension

Marina Stavrakantonaki, *University of Illinois at Chicago*

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

Allyson L. Holbrook, *University of Illinois at Chicago*

Young Ik Cho, *University of Wisconsin-Milwaukee*

Sharon Shavitt, *University of Illinois at Urbana/Champaign*

Saul Weiner, *University of Illinois at Chicago*

Noel Chavez, *University of Illinois at Chicago*

Prior research has shown that comprehension of survey questions is lower among members of minority racial and ethnic groups than among the dominant racial group (e.g., Holbrook, Cho, and Johnson, 2006). Intuition suggests that the ability to comprehend survey questions should be greater among first and second generation immigrants who are more acculturated to a host society than among more recent immigrants. There is little evidence available, however, to confirm this intuition. We report data from a survey conducted in Chicago with 300 Mexican American and Korean American respondents that employs behavior coding to identify comprehension difficulties among respondents. Using multi-level modeling and a set of 270 survey questions, we test the hypothesis that degree of acculturation is negatively associated with comprehension difficulties among respondents, controlling for respondent and question characteristics. Findings confirm our hypothesis. Implications for survey data collection with immigrant populations are discussed. Holbrook, A. L., Cho, Y. I., & Johnson, T. P. (2006). The Impact of Question and Respondent Characteristics on Comprehension and Mapping Difficulties. *Public Opinion Quarterly*, 70(4):565-595.

Is One More Reminder Worth It? If So, Pick Up the Phone: Findings From a Web Survey

Lisa Lin-Freeman, *IMPAQ International*

IMPAQ International conducted a web-based survey on behalf of the U.S. Department of Labor, Employment and Training Administration (ETA-TA). The purpose of the survey was to learn more about how training and technical assistance are being used in the public workforce system and the value that grant recipients place on the technical assistance they received. 400 workforce system grantees across the country were selected to participate. In total, 192 individuals completed the survey. Additionally, 20 submitted incomplete responses. We sent weekly email reminders throughout the field period. In total, five emails were sent from October 1 to November 4. Only one of the email reminders was followed by phone call. The cumulative number of completes increased after each reminder. However, with each reminder, we observed a declining number of incremental completes (after the first reminder, we received 44 completes, then 36 after the second reminder, and so on). We observed an exception to this trend when the reminder was followed by a phone call. After the phone call, we observed a substantial uptick in the number of completed surveys received that week. Specifically, the week prior to the reminder call, we received 14 surveys after the regular email reminder. After the reminder call, we received an additional 21. This study adds to the body of knowledge about the effectiveness of reminders in surveys. The decision to spend time, effort, and dollars on additional reminders is relevant, as survey researchers continue to balance project constraints and goals.

Representativeness of a Mixed-Mode Panel Across Time: Evidence from the GESIS Panel

Michael Bosnjak, *GESIS - Leibniz Institute for the Social Sciences*

Tobias Enderle, *GESIS - Leibniz Institute for the Social Sciences*

Klaus Pforr, *GESIS - Leibniz Institute for the Social Sciences*

In this presentation, we focus on the GESIS Panel, which offers the academic social science community a possibility to collect primary data free of charge within a probability-based, mixed-mode (online and mail survey) omnibus access panel encompassing almost 5000 panelists in early 2014. Data are being collected six times per year and are presumed to represent the attitudes, opinions and behaviors of the German population. Therefore, by relying on a random sample of the German general public, it is presumed that the GESIS Panel yields 'representative' data. However, such an assumption requires empirical evidence, providing answers to the following two core questions: (1) How 'representative' is the GESIS Panel at the outset, i.e. at the time of panel recruitment?; (2) How 'representative' is the GESIS Panel across time? In our presentation, we address both issues by first describing the construction of 'Representativeness-Indicators' (so-called R-Indicators; e.g., Schouten et al., 2012) for the GESIS Panel. This indicator is compared with classic balance indicators of auxiliary variables in gold standard samples. In the second part of our presentation, we map the developed R-Indicators across time, that is from recruitment to the most recent survey wave. In the third part, we relate the longitudinal variability in R-Indicators to known historical events and to changes in the methods and procedures employed to maintain the panel (e.g., incentive-schema changes, fielding period changes, administrative dropout of panelists). The approach we present may aid panel providers in developing their own sample quality monitoring system, survey methodologist may capitalize on our findings relating the variability of R-Indicators to specific events and interventions, and data users of the GESIS Panel will be provided with a fine-grained analysis of data quality in terms of representativeness.

Measures of Sleep: Methodologies, Potential Measurement Error and New Measurement Techniques

Heidi Guyer, *Survey Research Center, University of Michigan*

Short duration and poor quality sleep is associated with the onset and progression of many chronic diseases and affects cognitive, physical and emotional health as well. While many medical studies have focused on sleep apnea, non-apnea related poor quality sleep is of importance as well. With the rate of sleep disorders on the rise at all stages of the lifecourse, the implications of poor quality sleep are increasingly important. Given the increased risk associated with poor quality sleep, accurate measurement is important. Various methodologies and question formats have been utilized to measure duration and quality of sleep in survey research. This presentation will provide an overview of sleep measurement methodologies including question wording, prevalence rates, validation results and novel techniques in measuring the duration and quality of sleep. New technologies offer novel means of collecting additional data and may provide a means of validating survey responses as well. Mode and ease of administration of various measurement options, as well as associated costs, will be presented.

Comparing Manual and Automated Industry and Occupation Coding: Accuracy and Cost from the Perspective of the California Health Interview Survey

David Grant, *UCLA Center for Health Policy Research*

Royce Park, *UCLA Center for Health Policy Research*

Matthew Jans, *UCLA Center for Health Policy Research*

John Rauch, *Westat*

Marisol Frausto, *UCLA Center for Health Policy Research*

Assigning industry and occupation (IO) codes to open-ended employment responses can be a time-consuming, expensive, and error-prone endeavor. This paper compares manual coding with an automated, computer-assisted coding system developed by the National Institute for Occupational Safety and Health (NIOSH). California Health Interview Survey (CHIS) IO coding traditionally involves human coders that review and categorize respondent job titles based on verbatim text entries by CATI interviewers. The manual coding scheme uses 2010 Census occupation codes and the 2012 North American Industry Classification System (NAICS). It includes a double-blind process to validate IO codes followed by adjudication of conflicting codes. The NIOSH Industry and Occupation Computerized Coding System (NIOCCS) uses an automated coding algorithm to assign IO codes to text entries. A user can submit multiple records (batch-mode) via a Web interface (<http://wwwn.cdc.gov/niosh-nioccs/>). We randomly selected 1,000 manually-coded cases from 2013-2014 CHIS and processed them with the online NIOCCS system. Preliminary results suggest a clear benefit from using the NIOCCS as it substantially reduces the time and resources necessary to complete the coding, both in person-hours and project duration. Our final analysis will compare reliability of each coding system, and assess their success for industry and occupation codes separately. We will present the cost and data quality trade-offs of each system, as well as operational issues of implementation.

Top of the Mind or Cognitive Probing – Assessing Survey Data Quality in a Survey to Identify the Unmet Civil Legal Needs of Hard to Reach Low Income Households

Danna L. Moore, *Social and Economic Sciences Research Center*

Arina Gertseva, *Social and Economic Sciences Research Center*

Nathan Palmer, *Social and Economic Sciences Research Center*

The purpose of the survey was to identify and determine the incidence of unmet civil legal needs of hard to reach low income and otherwise vulnerable populations in Washington State. The sample consisted of an Address Based Random sample of low income census tracts (27% or more below poverty). The goals of this research were to test: 1) across survey modes, mail, telephone, and web the effectiveness of an eligibility screening question for targeting hard to reach low income individuals and households; and 2) the effectiveness of a top of the mind introductory question on civil legal problems for which they needed legal help as compared to the incidence of civil legal problems across 138 problem areas. The data collected provides information on effectiveness of the low income threshold screening question. This study uses indices to compare incidence rates for eligible households. This study demonstrates and provides insights into the measurement error comparing a top of the mind question as compared to incidence when very specific civil legal problem questions are used to cognitively trigger remembrance.

AAPOR 70th Annual Conference

Interview Quality in In-Depth Interviews: Findings from CDC's iQual Insight Study

Alisu Schoua-Glusberg, *Research Support Services*

Katherine Kenward, *Research Support Services*

Susan Berkowitz, *Impaq International*

Elizabeth Gall, *Impaq International*

James W. Carey, *Division of HIV/AIDS Prevention - CDC*

Quality aspects of qualitative research have received much focused attention over time, in particular, intra- and inter-coder reliability when analyzing qualitative data. In studies with large numbers of interviews and multiple interviewers, it is also very important to make sure each interviewer elicits all the data needed to answer the research questions, and in appropriate depth. With flexible instruments that include some scripted probes but also allow for spontaneous probing, it can be easy for the interviewer to miss asking certain questions, particularly if the respondent's narrative offers answers to questions that come later in the semi-structured instrument. The RSS Team (Research Support Services/Impaq/Emory University) conducted 100 qualitative interviews for the Insight study, a CDC-sponsored qualitative project (Division of HIV/AIDS Prevention) looking for factors that facilitate or interfere with care and treatment for HIV among HIV-positive Black and Latino men. Nine interviewers carried out the interviews in Los Angeles, Atlanta, Chicago, and Washington DC/Baltimore. We developed an assessment tool for judging interview fidelity. To evaluate how well each interviewer elicited answers to the questions of interest, we randomly selected 3 cases per interviewer. We then selected one set of 5 questions to evaluate for all interviews. We created a tool for this evaluation of interview quality, to allow us to assess how well the interviewer followed up with probes to elicit a rich, full narrative that addressed the research needs. The authors independently used the tool for the selected cases and met to discuss their ratings. This presentation will describe this procedure in detail and present findings about the usefulness of this quality assurance technique in qualitative research. We will also reflect on lessons learned, including recommendations for researchers managing multisite qualitative data collection under tight schedules, to keep the team working in sync to maximize fidelity and data quality.

Using Paradata to Predict Mobile Usage in Online Surveys

Kristin L. Cavallaro, *SSI*

As Smartphone ownership in the U.S. is expected to hit 75% by the end of 2014, it is critical to consider smartphones as a source of survey data collection. Previous research conducted by SSI has shown that 25% of 18-34 year old respondents are entering the survey via a smartphone. Given the difficulty of reaching this demographic group, it is critical to understand how to recruit and engage these respondents to provide the best coverage possible. But it is not just the younger age groups who are opting to use mobile devices; 14% of those attempting to take a survey on their smartphone were over the age of 55. In fact according to Nielsen, 51% of mobile users over the age of 55 own a smartphone. Given these statistics and the projection that these statistics promise, it is important to learn more about what drives people to take a survey on their smartphone and more importantly; what drives them away from taking surveys on their smartphone. Using paradata, we will examine the historical activities of respondents to identify the point where they start or stop using a smartphone to take surveys, and demonstrate the level of "crossover" between the modes: is respondent mode choice a function of time of day or survey length or topic, or other factors? This information will then be used to try to remedy any barriers to taking surveys on a smartphone while building reach and respondent engagement. Variables that will be considered include the device the respondent joined on, the

AAPOR 70th Annual Conference

completion status of their first survey on a smartphone, demographics, project information and more.

The Effect of Providing Think-Aloud Examples and Practice on Cognitive Interviewing in Nepal

Kristen Cibelli Hibben, *University of Michigan*

Jennifer Kelley, *University of Michigan*

Ting Yan, *Westat, The Joint Program in Survey Methodology*

Cognitive interviewing is most effective when respondents verbalize their thought process in response to the survey questions and interviewer probes. This process may be unfamiliar and difficult for some respondents, particularly in contexts outside of the United States and Europe where there is little survey tradition. Little guidance exists in the literature on how to apply cognitive interviewing in such contexts. This paper examines the effect of providing respondents with examples and practice think-alouds before the main cognitive interview. The cognitive interviewing study was carried out in Nepal where experience with surveys is rare. The study tested questions for an international household survey on disability. The Institute for Social and Environmental Research (ISER) in Chitwan Valley, Nepal and the University of Michigan conducted the cognitive interviews (n=40). Respondents were randomly assigned to a treatment or a control condition. In the treatment condition, interviewers demonstrated thinking aloud in response to an example question and then asked the respondent to practice thinking aloud to two additional practice questions. Respondents in the control condition did not receive examples or practice questions. Interviewers provided an introduction explaining the purpose of the interview and the cognitive interviewing process to respondents in both conditions. All respondents were asked debriefing questions on their cognitive interviewing experience (e.g. how easy/difficult to think aloud; how comfortable/uncomfortable). This paper examines the effect of providing examples and practice think-alouds on the ability to capture sufficient data for qualitative analysis. Specifically, we analyze respondents' answers (e.g., the number of mentions unprompted, the number of codable responses, the amount of information learned) and the results of the debriefing questions. The paper concludes with lessons-learned and suggestions for future research on the application of cognitive interviewing in cultural contexts outside of the United States and Europe.

Patterns of Nonresponse to Health, Diet and Exercise Measures Conditioned on Body Mass Index

Rossi Dobrikova, *Experian Marketing Services*

Josephine Leonard, *Experian Marketing Services*

Jennifer Bouterse, *Experian Marketing Services*

Historically, survey researchers have been interested in collecting data on attitudes towards various health issues, particularly as they relate to diet, nutrition and exercise. One of the potentially important contributing factors in facilitating the collection of these attitudes and behaviors may be the actual height and weight of the respondent themselves. A more complex metric along these lines is Body Mass Index (BMI) – a simple measure of body fat that incorporates a person's height and weight. BMI correlates well with more precise laboratory measures of body fat and is often used as one indicator of an individual's well-being. One hypothesis that is likely worth investigating is the effect of deviations from normal standards for BMI has on the item nonresponse rates for specific diet, nutrition and exercise measures in a survey. Individuals who are classified as overweight or obese given their BMI metric may also be more likely to fail to respond to questions concerning diet, nutrition, exercise and even health

AAPOR 70th Annual Conference

care attitudes and behaviors due to the stigma of occupying one of these statuses. In this paper we first explore the difficulties in extracting measures for height and weight in a survey and discuss a strategy that has proved successful in attenuating nonresponse to these sensitive questions. We then turn to examine the potential relationship between respondent BMI and nonresponse to survey measures of diet, nutrition, exercise and attitudes towards healthcare and healthcare providers. A multivariate statistical model of item nonresponse is built for a number of these measures that include demographic controls as well as the main independent variable of BMI. The data comes from a large, national probability sample of approximately 25,000 respondents.

Using Pre-Incentives to Increase BRFSS Response Rates

Piper Jean DuBray, *ICF International*

Jessie Hammond, *Vermont Department of Health*

Survey response rates have greatly declined in the past decade, compelling researchers to seek new ways to increase participation. The Vermont Department of Health (VT DOH) and ICF International (ICF) conducted a pilot study in 2014 for which we included a \$2 cash pre-incentive with the advance letter for a portion of the landline sample, to determine 1) whether including a \$2 cash pre-incentive increases the response rate, and 2) whether the pre-incentive is cost-effective. For the 2014 VT BRFSS, approximately 23% of the sample each month has a matched address and is sent an advance letter. ICF randomly divided the monthly sample with matched addresses into 2 groups: sample in one group received a \$2 pre-incentive with the advance letter; sample in the other group received only the advance letter (no pre-incentive). We analyzed data for three groups: sample sent a letter which included a pre-incentive; sample sent a letter only, not including a pre-incentive; and sample that was not sent a letter. Results based on data for the first six months of 2014 indicate that including a pre-incentive does significantly increase the response rate. Including pre-incentives with half of the letters increases the statewide AAPOR #4 response rate by 3.3 percentage points. Analyses determined that the incentives are cost-effective. We compared the cost to increase the sample size to the cost of administering the incentives, and determined that 69% of the costs associated with the incentive pays for itself in additional interviews. Preliminary results also demonstrate that including a pre-incentive does not increase response rates for particular demographic groups. Finally, including a pre-incentive does not reduce the number of calling attempts needed to obtain a completed interview or otherwise resolve the record. We will conduct further analyses for the full 2014 survey year and present the annual findings.

Modeling Uninsurance Estimates at the County Level

Joanna Turner, *University of Minnesota, SHADAC*

Peter Graven, *Oregon Health & Science University*

Brett Fried, *University of Minnesota, SHADAC*

Donna Spencer, *University of Minnesota, SHADAC*

Kathleen Call, *University of Minnesota, SHADAC*

With the implementation and evaluation of the Patient Protection and Affordable Care Act (ACA) there is increasing demand for timelier estimates of health insurance coverage, as well as estimates for sub-populations and sub-state geographies. Federal programs, such as the American Community Survey (ACS) and the Small Area Health Insurance Estimates (SAHIE) provide uninsurance at the county-level and several states also conduct their own surveys. Individual state surveys can ask more detailed questions focused on particular situations and programs of their state, but they often lack sample size to produce reliable county-level

AAPOR 70th Annual Conference

estimates. We conducted a case study to develop county-level estimates from a state-specific survey using small area estimation techniques. Building on the rich set of tools available for producing small area estimates we developed an innovative model that combines multiple sources of outcome data and uses auxiliary data to improve the precision of the estimates. We modeled direct estimates of uninsurance at the county-level from a state survey with covariates from the ACS and administrative records. We then combined the modeled estimates with SAHIE estimates. The two-step approach was used to evaluate the precision of the estimates at each step. While it is appealing to use only the state-specific survey and covariates, the predicted county-level estimates have high variance. Combining the modeled estimates with SAHIE reduces the variance, but given the precision of SAHIE, the combined estimates are very similar to SAHIE. While our approach can be used for other outcomes, given the availability and precision of sub-state uninsurance estimates from ACS and SAHIE, modeling uninsurance estimates from state-specific surveys may not yield much gain. Having a benchmark was unique to our study, and as small area estimation techniques become more prevalent for evaluating health policy it is important to provide transparency including measures of uncertainty and detailed documentation.

Friday, May 15
4:15 p.m. – 5:45 p.m.
Concurrent Session F

Mini-Conference: Says Who? Effects of Survey Sampling Methods on Respondent Characteristics and Survey Findings

Mini-Conference: Says Who? Effects of Survey Sampling Methods on Respondent Characteristics and Survey Findings

Jennifer Hunter Childs, *U.S. Census Bureau*
Elizabeth M. Nichols, *U.S. Census Bureau*
Kathleen Kephart, *U.S. Census Bureau*

With the current discussion of relative strengths and weaknesses of non-probability panels, this panel highlights papers investigating differences between nonprobability and probability panel findings, both in private industry and in the federal government. In particular, we are interested in the respondents of probability versus non-probability samples and how recruitment methods for both types of pools influence findings. We also address the concept of “fit for purpose” with regard to these different types of recruiting and sample pools.

A Census-Affinity Nonprobability Panel: Uptake and Characteristics of the Participants

Jennifer Hunter Childs, *U.S. Census Bureau*
Elizabeth M. Nichols, *U.S. Census Bureau*
Kathleen Kephart, *U.S. Census Bureau*

Since March 2013, the U.S. Census Bureau has offered the ability for the public to sign up to participate in online research studies. The sign up is through the Census Bureau website. There

AAPOR 70th Annual Conference

is no monetary incentive associated with signing up and there has been no publicity thus far about the panel. Currently, over 15,000 people have signed up to participate with approximately 1,000 new emails joining each month. The Census Bureau has run six studies with this panel, investigating email invitations to a survey. Response rates to the surveys have hovered between 10 to 20 percent, relatively high for a nonprobability panel. Survey data reveal these participants tend to be older, more educated and more likely to be in the workforce than the general U.S. population.

Building a Quality Nonprobability Panel: Methods, Problems and Doing It All Innovatively

Annie Pettit, *Peanut Labs*

The vast majority of online research panels create a website and find ways to attract people to the website and subsequently register for the panel. The Peanut Labs panel works differently in that panelists don't have to be aware of or search out online panels. Rather, panelists are presented with surveys while they are in an originating website – whether it be a popular social media website, a gaming website, a loyalty program, or some other website where points and credits are available. This session will describe the advantages and disadvantages of this unique method, and discuss a number of the processes used for generating quality data that mirrors census representative data.

Non-Probability Surveys Online: Does "Empanelment" Affect Results?

Sarah Cho, *SurveyMonkey*

Noble Kuriakose, *SurveyMonkey*

Online surveys are typically conducted using a recruited panel or an ad hoc flow of web survey takers. SurveyMonkey has the capacity to do both, and this paper uses a unique set of data to explore what differences, if any, stem from respondents' agreement to participate in panel research. Nearly three million survey responses are collected each day on the SurveyMonkey platform, and upon completion of these user-created surveys, SurveyMonkey conducts research surveys from this ad hoc flow of respondents. These millions of user-generated surveys also serve as the recruitment source for SurveyMonkey Audience, a non-probability based panel that incentivizes its members by offering to make small, charitable donations for taking surveys. To date, over 4 million respondents have signed up to participate in this online non-probability panel. In this paper, we will compare surveys conducted using the recruited panel to those conducted among the broader set of ad hoc survey takers. We investigate the similarities and the differences between the samples, both in terms of demographics and opinions on a wide variety of issues.

Context Clues: The Impact of Questionnaire Content on Self-Selection Bias and Context Effects in a Probability-Based Sample

Casey Eggleston, *U.S. Census Bureau*

Since 2012, the Census Bureau has partnered with the Gallup Organization to collect nightly survey responses measuring U.S. adults' trust in the Federal Statistical System (FSS). Through March 2014, these trust items were included in the Gallup-Healthways Wellbeing Index, one of Gallup's two daily tracking surveys. In September 2014, due to time constraints on the Wellbeing Index, the FSS Public Opinion Survey was migrated to Gallup's daily Political and Economic survey track. Aware that this change in instrument could impact both respondent self-

selection into the survey and responses to individual questionnaire items due to context effects, Census elected to include several weeks of simultaneous data collection on both of Gallup's daily tracking surveys. The current paper compares survey responses across the two topical tracks to examine the effect of questionnaire content on demographics of survey respondents as well as item-level responses to the FSS Public Opinion Survey.

Selected or Self-Selected? Part 1: A Comparison of Methods for Reducing the Impact of Self-Selection Biases from Non-Probability Surveys

David J. Dutwin, *SSRS*

Trent D. Buskirk, *Marketing Systems Group*

Non-probability based opt-in samples and panels are beginning to emerge in popularity in many research areas including marketing, consumer and social and political sciences. Some research has been devoted to exploring methods for reducing the impact of self-selection bias on the overall estimates derived from non-probability samples including simple post-stratification calibration, propensity score adjustments and sample matching that links non-probability respondents to respondents from a temporally relevant probability sample. The efficacy of any of these adjustment techniques lies on the ability to match or link respondents across surveys and on the availability of variables that are related to panel participation. In this study we will apply three different adjustment techniques using a probability and non-probability samples and evaluate the resulting biases in the final estimates of selected cross classified demographic variables. Specifically, we will apply adjustments based on: (1) direct post-stratification weighting calibration; (2) propensity score adjustments and (3) sample matching using data collected for a series of studies about sports engagement and media consumption using both non-probability samples as well as a concurrent RDD dual frame and ABS probability samples. The post-stratification method will calibrate non-probability respondents to standard demographic marginal distributions using standard iterative proportional fitting routines while the propensity method will create a direct adjustment factor using a model predicting panel type from a host of demographic, psychographic and behavioral variables. The matching approach will be based on linking each probability respondent to a non-probability respondent who has the closest value of the propensity score. We also present the overall raw distributions of respondents from the two panels on a host of demographic outcomes to further explore the tenability of the assumptions required for the matching and propensity adjustment methods.

Selected or Self-Selected? Part 2: Exploring Non-Probability and Probability Samples from Response Propensities to Participant Profiles to Outcome Distributions

Trent D. Buskirk, *Marketing Systems Group*

David J. Dutwin, *SSRS*

Methods for calibrating and linking non-probability panels to probability samples are beginning to emerge in order to reduce the potential impact of self-selection biases in resulting estimates based on opt-in samples. Covariates used in generating propensities/matches must be available for respondents from both samples and variables used to link respondents across the two samples should have the same basic ranges. While comparisons of differences in estimates derived from matched samples and probability samples has been previously explored, what remains to be seen is whether or not the distributions for outcomes of interest from the matched non-probability sample mimics that of the probability sample. In particular, are values obtained from non-probability samples (linked or otherwise) more likely to have higher intra class correlation coefficients that one would expect from clustered probability samples. In this

AAPOR 70th Annual Conference

research we extend the evaluation of differences in estimates to examining differences in respondent profiles and outcome variable distributions using probability and non-probability samples obtained for studies of media consumption and sports engagement. In particular, we apply metrics of network complexity from social network analysis as well as dissimilarity indices derived from random forest models to the demographic profiles from both samples to understand how similar the respondents are across the two samples. We will also compare the intra-class correlation coefficients for various substantive outcomes across the two samples. We hypothesize that the self-selection mechanisms associated with non-probability samples may imply larger cluster effects, more similarity and tighter network structure compared to outcomes from probability samples. Finally, we examine differences in outcome variable distributions across the probability sample and a matched non-probability subsample created by linking probability respondents to non-probability respondents using both a propensity model based on demographic characteristics as well as a similarity index derived from a random forest model.

Novel Uses of Social Media for Survey Application

How We Data Mined Social Media to Predict the World's Largest Music Poll

David Quach, *The Nielsen Company*

The radio station "Triple J" runs a yearly countdown of the 100 top songs of that year as voted by its listeners. Each year about 1.5 million votes are submitted from listeners in 170 different countries—making this countdown of songs, named the "Hottest 100", the largest music poll in the world. Each year, many people try to predict which songs will make the countdown, including which song will be number 1. This year, listeners were posting photos of their voting confirmation emails on Instagram and Twitter. These photos contained the list of songs that the listener had voted for. We downloaded these images and used Optical Character Recognition (OCR) software and manual methods to count almost 18,000 votes. Using this data, we correctly predicted 85 of 100 songs that were on the Hottest 100 of 2013, including the number 1 song. Our prediction was posted to the following website: www.Warmest100.com.au.

Linking Individuals' Twitter Data with Survey Data: Challenges and Approaches

Ellen Wagner, *University of Michigan*

Josh Pasek, *University of Michigan*

Darren Stevenson, *University of Michigan*

Researchers are increasingly interested in the extent to which social media data, particularly from Twitter, might be able to yield conclusions about society that were traditionally the domain of surveys. And a number of recent studies suggest that Twitter and survey data sometimes correspond. But the value of Twitter data analyses depends on whether Twitter data tell us about individual Twitter users or instead provide only an aggregate-level picture. To understand this, we need to be able to link Twitter data with survey data from the same individuals. This study explores the challenges and opportunities associated with two methods of linking these two types of data: requesting Twitter handles from survey respondents and placing Twitter advertisements to recruit individuals to surveys. We report some success in using both methods and show how each can lead to different types of inferences about the connections between these two types of data. We discuss the benefits and drawbacks of both approaches and identify different circumstances for which each of these methods can be used.

AAPOR 70th Annual Conference

Putting Social Media Data in a Probability Frame Context

Edward Paul Johnson, *Survey Sampling International*

Lisa M. Drew, *Two42 Solutions*

Mohammad Hamid, *Two42 Solutions*

Over the years, the communication landscape has dramatically changed. Social media, among other forms of new media, has not only changed the way that individuals consume information, but also how public opinion is shaped. In analyzing social media data researchers face numerous challenges related to audience sampling, legal compliance, and individual privacy which can introduce bias when applying the findings to the general population. Our study on the integration of telephone survey methodologies and social media sheds some light on ways to minimize these challenges, while still maximizing research utility. SSI and Two42 Solutions present two methods for structuring social media data inside the framework of a probability sample: (1) matching telephone survey responses to social media accounts at the individual level - directly or through a nearest neighbor search and (2) matching aggregated issue-based questionnaire responses to weighted social media content. We draw a random sample from the Michigan voter file to provide a probability sample frame. We then conduct a short telephone survey that includes an opt-in to be tracked on social media afterwards. The individual matching method invites those that opted-in (and nearest neighbors of those that did not opt in) and pushes a notification to them for permission to track. The social media analysis is then constrained to these individuals. The aggregate issue-based method looks at all the social media data, codes it into topics and then weights the topics according to importance as seen in the telephone survey. Additional analysis of the individual matching technique includes a comparison of respondents' attitudes and behaviors as conveyed through telephone survey responses and social media data. Identification of the similarities and differences based on the two modalities include: top issues, polarization, demographic and psychographic characteristics.

Can We Leverage Facebook's Social Structures for Survey Recruitment?

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

Rupa S. Valdez, *University of Virginia Department of Public Health Sciences*

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

Ishan C. Williams, *University of Virginia School of Nursing*

Obtaining access to representative samples for health-related research can be costly and time consuming. The 2014 AAPOR Task Force report on social media identifies study recruitment as one of the main potential uses for social media in survey research. In 2008-09, Christine Bhutta (2012) was able to recruit over 7,500 baptized U.S. Catholics in just one month using Facebook. Bhutta created Facebook groups and then recruited individuals into those groups, allowing her to access them directly as potential survey respondents. In contrast, recent studies that have used Facebook advertisements for recruitment into health studies report mixed results. The current AHRQ-funded study assesses the current feasibility, benefits, and challenges of recruiting for health surveys by leveraging the social structures embedded in the Facebook platform, using a method similar to Bhutta's. Our recruitment strategy targeted Facebook users with Type 2 diabetes and involved creating a study-related Facebook group and asking administrators of other groups and pages to publicize our group to their members. Group members were then directly invited to participate in two online surveys: a pre-study demographic screener and a substantive survey about how they communicate health information. To assess the recruitment method, we look at recruitment rate, cost, content of interaction, and characteristics of the sample obtained. Our study has not achieved the level of recruitment success that Bhutta achieved when Facebook was novel; we have enrolled fewer

AAPOR 70th Annual Conference

than 200 persons with diabetes. The content of our interactions with group managers and diabetes patients has raised interesting ethical issues concerning suitable target communities and appropriate boundaries between researchers and participants. Leveraging the social structures of Facebook for health-related research is currently feasible for obtaining small samples appropriate for pilot or qualitative research. But, similarly to an online panel, the process requires a substantial and continuous investment in respondent relations management.

How Important is the Visual?: Differences in Click-Through Rates for Social Media Ads Using Various Visuals

Ashley M. Schaad, *ICF International*

Rikki Welch, *ICF International*

This paper focuses on a highly targeted pay-per-click ad campaign launched through Facebook. The campaign was designed to recruit focus group participants from Greater Houston fitting one or more of the following demographic categories as part of a disaster preparedness research project: individuals with ambulatory or transportation difficulties, seniors 65 and older living independently, caregivers for seniors or individuals with cognitive impairments, individuals who are deaf or hard of hearing, and individuals who are blind or have low vision. The ads were posted in English and Spanish, with separate groups of ads developed for each of the demographic populations. We used up to four different images associated with each ad text, monitoring the click-through rate to determine which image best captured the audience's attention (i.e., led to the highest percentage of clicks on the ad). The text for the ads within each ad group remained the same. The images included a dollar sign to focus on the paid nature of the focus groups; a hurricane to highlight the focus group topic of disaster preparedness; the state flag of Texas to incorporate a familiar image likely to resonate with Houston residents; and a photo representing the demographic group of interest, such as a Hispanic senior or a handicap parking sign. The image garnering the highest click-through rate within each ad group varied, with the dollar sign and Texas flag each garnering the highest click-through rate within four of the ad groups and the hurricane garnering the highest rate within the remaining two ad groups. This study illustrates the importance of using multiple images with online ads and suggests that familiar images, such as a state flag, may be equally effective as images related to the actual data collection effort.

Building Better Election Polls

It's My Party: Comparing Actual Party Registration, Self-Reported Registration and Self-Identification

Steven M. Koczela, *The MassINC Polling Group*

Richard Parr, *The MassINC Polling Group*

Public polls and poll reporting use a variety of descriptors of party affiliation. Most common is party self-identification: asking respondents with which party they most identify. Other methods include self-reported party registration and individual party registration drawn from lists of registered voters. The implications of each method are not well-understood. Using data from five waves of the WBUR Tracking Poll on the 2014 Massachusetts General Election, we explore the differences among these three methods of identifying voters' political affiliations. Our findings show people frequently misreport their own party registration. This suggests self-reported party registration is not a reliable substitute for party registration information drawn

AAPOR 70th Annual Conference

from lists of registered voters. We also examine party self-identification versus actual party registration, to show that the groups of voters who self-identify as Democrats or Republicans overlap only partially with party groupings from official voter rolls. Those who report registration with a party or self-identify with a party differ systematically from registered members of the party on key opinion dynamics. We examine differences in vote preference and favorability for the Democratic and Republican candidates in the governor's race among actual registered Democrats and Republicans, those who self-reported being registered to either party, and those who self-identified with each party. Key differences in the degree of partisanship in each of these groups are explored. In three of the five waves of this weekly poll, respondents were asked with which party they were registered. In the other two waves they were asked with which party they most identified. Responses were then compared to respondents' actual party registration, as recorded in our sample source, the Massachusetts voter file. These results are significant given the many ways of discussing partisanship when reporting election poll results.

Again and Again and Again: Dynamic Bayesian Forecasting of Election Polling Data

Robert A. Petrin, *Ipsos Public Affairs*

Alan Roshwalb, *Ipsos Public Affairs*

Neale El-Dash, *Sleek Data*

Marcus Maher, *Ipsos Public Affairs*

Election night race-calling requires assimilating multiple sources of vote count information, historical data, and exit polls. Calling elections is further complicated by the fact that polling data can be reported erratically and in a non-representative manner throughout the evening, while the data that do roll in can be subject to error and frequent updating. These challenges must be reconciled with the news-value of election results which requires balancing speed with accuracy, as well as the need to forecast election-outcomes in a principled and transparent manner. Bayesian methods (Carlin and Louis, 2008; Gelman et al., 2013; Gelman and Hill, 2006) offer a flexible and robust framework for addressing these challenges. Accordingly, this paper explores the application of Bayesian methods to calling races, drawing upon Ipsos's experience with Thomson Reuters from the 2014 U.S. Midterm Elections. The specific focus of this paper is on Bayesian methods for dynamically updating priors to incorporate election-night turnout and trending in toss up races. Specific research questions this paper will address are: What are some effective ways of updating model priors in a dynamic and fluid analytic environment?, What is the optimal rate at which to shrink posterior parameter estimates away from model priors?, Are there robust rules of thumb (for example, number of precincts reporting) that can be applied to establish when races can be optimally called? The data used for this study are taken from the live data-feed collected by Ipsos for Thomson Reuters's U.S. Election Services program. The performance of each of the proposed Bayesian methods are evaluated using key battleground or "toss up" races from November 4, 2014, applying Jackman's (2005) approach for assessing bias in political polls. The paper concludes with caveats and recommendations for future election forecasting and race-calling efforts.

Model Based Inference Using RBS Survey, Machine Learning and Voter List Masahiko Aida, *Civis Analytics*

Author expands the previously discussed advantages of model based inference in political horse race estimates (Aida, 2014). In addition to the Presidential polling data of 2012, author adds 2014 Senatorial polling data from midterm election to test the validity of the approach across the different contexts. One of the advantages of model based inference (compared to

AAPOR 70th Annual Conference

design based approach) is that it provides stronger defense against non-response and non-coverage biases. Whereas traditional weighting adjustment will struggle with variance-bias trade off, model based correction will reduce greater amount of bias without sacrificing precision, provided model is well specified and sample size being large. Another benefit of this approach is it can bypass infamously inaccurate likely voter screen (Rogers & Aida, 2014) that is known to include various reporting errors. This is done by replacing self-reported likely voter screen with turnout score made of past vote histories. Once survey is conducted and responses matched to voter list, machine learning is used to create predictive models of candidate support using various fields from voter list. Lastly, population estimates are made by aggregating support scores from predictive model and turnout score. Author compares actual results of 2012 and 2014 against modeled estimates for each race to evaluate accuracy of the estimates, then discuss strength, weakness and challenges of model based inferences.

Does Candidate Order Matter? Impact of Matching Ballot Order on Pre-Election Poll Accuracy

Scott Clement, *The Washington Post and University of Maryland Joint Program in Survey Methodology*

Jon A. Krosnick, *Stanford University*

Recent studies of election ballot order effects find voters tend to pick candidates listed first, while telephone poll respondents tend to pick the last candidate they hear. Candidates in U.S. elections are listed on ballots in a consistent order throughout most states; as such, ballot order effects have a potential to impact election outcomes as well as pre-election poll accuracy. While research on ballot order effects has grown, no known literature has explored strategies for pre-election polls to mitigate countervailing candidate recency order effects in polls and primacy effects on the ballot. This paper assesses the accuracy of two strategies for ordering candidates in pre-election polls – (1) matching the order of candidates listed in a poll to the actual ballot order and (2) listing candidates in the opposite order of the ballot. These two approaches are compared to the common polling practice of randomly assigning all candidates to one of these two orders. To test these approaches I analyze a national probability sample survey in the asking more than 6,000 likely voters presidential vote preference in the final weeks before the 2012 election. Respondents were randomly assigned to hear Barack Obama or Mitt Romney first in the presidential vote choice question. The actual local ballot order for most respondents was identified using state and local election laws. As such, the poll mimics an experiment where each respondent was randomly assigned to hear candidates in the same or opposite order of their own local ballot. I compare the accuracy of these candidate-ordering approaches in estimating national vote support.

Early Voting Effects on Pre-Election Poll Estimates

Michael P. McDonald, *University of Florida*

Michael Martinez, *University of Florida*

Christopher McCarty, *University of Florida*

Daniel Smith, *University of Florida*

Survey organizations are typically rated by how well their pre-election polls forecast the election outcome. This approach is attractive because the poll estimate can be compared to the truth, which is unavailable for many other attitudinal and behavioral measures that polls estimate. There are, however, sub-samples of pre-election poll forecasts that can also be validated against the truth that heretofore have not been systematically analyzed. In states with a high usage of early voting, pollsters often ask respondents if they have already voted and report

AAPOR 70th Annual Conference

candidate support estimates for respondents who report that have already voted and those respondents who meet the likely voter screen. Election officials in states such as Florida, Iowa and North Carolina also report their election results by method of voting. We use these data to validate pre-election poll candidate support levels for surveys that report candidates' support by method of voting. A further analysis of a registration based sample of respondents to three 2014 University of Florida pre-election polls, where individual level respondents' vote method are validated, provides further insights to respondents' behavior that may challenge pre-election polls' estimates for early and Election Day voters. We analyze which respondents are more likely to misreport their voting method and who are more likely to refuse candidate support questions if they report already voting.

The Timeline of Elections: A Comparative Perspective

Christopher Wlezien, *University of Texas at Austin*

Will Jennings, *University of Southampton*

Scholars are only beginning to understand the evolution of electoral sentiment over time. How do preferences come into focus over the electoral cycle in different countries? Do they evolve in patterned ways? Does the evolution vary across countries? This paper addresses these issues. We consider differences in political institutions and how they might impact voter preferences over the course of the election cycle. We then outline an empirical analysis relating support for parties or candidates in pre-election polls to their final vote. The analysis relies on over 26,000 vote intention polls in 45 countries since 1942, covering 312 discrete electoral cycles. Our results indicate that early polls contain substantial information about the final result but that they become increasingly informative over the election cycle. Although the degree to which this is true varies across countries in understandable ways given differences in political institutions, the pattern is strikingly general.

Reducing "Grid"-Lock: Advances in Formatting Questions in Grids Versus Alternative Formats

Gridlocked: The Impact of Adapting Survey Grids for Smartphones

Ashley Richards, *RTI International*

Rebecca Powell, *RTI International*

Joe Murphy, *RTI International*

Shengchao Yu, *New York City Department of Health and Mental Hygiene*

Mai Nguyen, *RTI International*

Paper and web surveys often include grid-style questions formatted to save space and avoid repetition. However, this format is discouraged because it can result in high item nonresponse rates and straight-lining. Even more importantly in today's survey environment, grids often do not display correctly on devices with small screens, like smartphones. This is increasingly concerning as the percentage of web surveys taken on mobile devices continues to rise. The World Trade Center Health Registry has used grids to conserve space in three waves of surveys. To maintain comparability with prior waves and between paper and web modes, we included several grid items in the 4th wave of the survey, launched recently to a sample of 68,000 Registry members, to ascertain health conditions. Due to the rising number of respondents using smartphones to complete web surveys, we used responsive web design programming to automatically reformat grids into a series of individual items when a small

AAPOR 70th Annual Conference

screen device was detected. This method allowed us to retain grids used in previous years, yet address the issue of grids displaying poorly on mobile devices. In this paper we compare responses to grid items implemented on larger-screen devices to the corresponding individual questions on smartphones. More specifically, we compare rates of reported health conditions by question format. Because respondents reported the same health conditions and year of diagnosis in previous waves of the survey, we are able to compare consistency with previous waves and rates of new conditions reported, by format. In addition, we compare rates of missing data and straight-lining to see whether the smartphone-optimized versions suggest poorer, equal, or better data quality than the grids. The findings will be useful for guiding the design of future surveys addressing smartphone administration, but for which removing grids entirely may not be the best solution.

Grids and Online Panels: A Comparison of Device Type from a Survey Quality Perspective

Mengyang Wang, *University of Nebraska–Lincoln*

Allan McCutcheon, *University of Nebraska–Lincoln Survey Research and Methodology Program*

Laura Allen, *University of Nebraska–Lincoln*

Due to the widespread adoption of mobile devices, and the increased share of Internet traffic from mobile devices, these can now feasibly be applied to conduct surveys. However, mobile web surveys have several unique features that may adversely affect data quality (e.g., smaller screen and keyboard size). The increased navigation burden potentially makes respondents more likely to satisfice when responding on mobile devices than those responding on a computer. This research focuses on an experiment designed to explore which satisficing behaviors differ when responding to grid questions depending on: 1) the device type respondents used; 2) the location of the grid within the survey questionnaire; and 3) the interaction effect between grids location and device type. To investigate these research questions, 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, as well as across device types. Survival analysis and ANCOVA are used to explain the plausibility of survey satisficing regarding question location of grid questions and device type. Specifically, satisficing behavior is measured in terms of several indicators including (1) speed-up, (2) breakoffs, (3) no-opinion responses, and (4) straight-lining. Evidence from our previous studies indicates smartphone users are more likely to have higher rates of speed-up, breakoffs and no-opinion responses, demonstrating more satisficing among smartphone users. Moreover, our previous studies indicate that as the survey progresses, respondents who opt for “don’t know” responses spend less time responding, indicating satisficing behavior to increase with use of “don’t know” option later within grids in the survey.

The Effects of Grids on Web Surveys Completed with Mobile Devices

David Sterrett, *NORC at the University of Chicago*

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

Gwendolyn Rugg, *Cultural Policy Center at the University of Chicago*

Ethan Raker, *Columbia University*

Jiwon Baek, *Yonsei University*

Bilgen Ipek, *NORC at the University of Chicago*

Question grids are common on web surveys, and studies show that grids can affect how respondents complete surveys. However, there is little research that investigates the effects of grids on web surveys completed on mobile devices. In this paper, we evaluate the effects of question grids on response quality and measurement error for surveys taken on phones or tablets. Our study draws on a probabilistic web survey. The survey included an experiment in which respondents were assigned to one of three question format conditions: one large grid, two small grids, or single item per page. We analyze how question grids affect response times and nondifferentiation as well as explore the interaction effects between grids and devices. Reductions in time associated with question grids were greater for surveys completed on mobile devices as opposed to those completed on computers. Likewise, the increases in nondifferentiation associated with question grids were greater for surveys completed on mobile devices. We find that effects of question grids on responses in web surveys can differ across devices, and researchers should be cautious of using grids on web surveys as more people opt to do surveys on phones or tablets.

Beyond the Yes-No Grid: Expanding Consideration of Events Leads to Higher Endorsement

Randall K. Thomas, *GfK Custom Research*

Frances M. Barlas, *GfK Custom Research*

Multiple response formats (MRF – ‘select all’) have been shown to yield lower endorsement rates compared to dichotomous yes-no grids (YNG). Some studies have shown that other response formats can yield even higher rates of endorsement. As an example, self-reported disability is higher for the YNG than for MRF, but the use of a combination grid (CG) that assesses household-wide occurrence yields the highest level of self-reported disability. One oft-debated theory to explain the higher levels of endorsement for the grids is that they lead participants to consider less salient events. In this study, we compared the MRF and YNG with a CG for past 6 month purchase behavior. We randomly assigned respondents to response format and asked about the purchase of 10 products, including men’s clothing, groceries, and sporting goods. We used the Advertising Research Foundation’s FOQ2 dataset, which had 57,104 completes from 17 different opt-in sample sources (each providing at least 3,000 completes) and fielded in January 2013. Respondents to the MRF and YNG were asked to consider purchases both online and in brick-and-mortar stores. However, the CG used separate responses of ‘Yes, physical store only, Yes, online only, Yes, both physical store and online, No, did not purchase.’ For items reported as having been purchased, we asked about the recency of the purchase. We found that the CG took longer to complete than the YNG, which took longer to complete than the MRF. The CG led to significantly higher levels of endorsement than YNG, and YNG was significantly higher than MRF. Supporting the salience explanation for the differences in response formats, we found that respondents indicated purchase of products more recently with the MRF and less recently with the CG. Based on findings for this and other studies, we provide recommendations for measuring incidence.

AAPOR 70th Annual Conference

Using Grids Versus Mark All that Apply in a 2014 National Mail Survey

Danielle Battle, *American Institutes for Research*

Brett Bejcek, *The Ohio State University*

Mahi Megra, *American Institutes for Research*

Self-administered questionnaires rely on a carefully planned design in order to attain high response rates and increased data quality. During the design process, survey designers must create a clear organizational structure that allows the respondent to complete the survey as intended (Fanning, 2005). However, survey designers are balancing two important factors, first is the need for parsimony by reducing space to save on cost, but they also need to create an organizational structure that is intuitive to respondents and will yield high quality data.

Specifically, grid organizational systems are one type of system that allow for differing questions with the same set of criteria to be answered in a more condensed space than a question by question format. Because grids can vary in complexity, there are a wide variety of challenges associated with their use. For example, a grid with a simple rating system can be prone to higher levels of straight-lining, whereas a complex grid might have higher item non-response rates. As part of the 2014 National Household Education Surveys Feasibility Study (NHES-FS), a self-administered mail-based survey, a split panel experiment on a set of four questions was conducted to determine whether a grid system would or would not be feasible for future administrations. Half of the sample received a form with the grid rating system, and the other half received a form with similar questions in a check all that apply and rank order format. The assessment of common problems related to grids versus a non-grid mark all that apply format will allow for better utilization of organizational structures in future surveys by knowing which format provides the highest quality data. By comparing unit response rates, distribution across items, item non-response rates, and straight-lining patterns, analysis will show which organizational structure is better suited for self-administered mail-based surveys.

Sensitivity, Privacy and IRBs

Overzealous Institutional Review Boards vs. Frustrated Academic Researchers

Don A. Dillman, *Washington State University*

On a nearly daily basis conflicts arises between college and university Institutional Review Boards (IRB's) charged with protecting human research subjects and the scientists who submit survey plans for IRB approval. Differences between these groups on how surveys should and should not be implemented often result in an uncomfortable dance between study investigators and IRBs. Researchers tend to focus on improving response rates and reducing nonresponse error; while IRBs focus on ways that respondents should be protected, which sometimes require procedural changes that appear to have little if anything to do with human protection, but may affect the willingness of individuals to respond to survey requests. As a result, survey investigators often find themselves in a difficult dilemma. On the one hand, they would like to contest an IRB decision when approval is withheld for reasons having more to do with survey procedures they see as beyond the purview of IRB responsibilities. On the other hand, questioning IRB decisions may slow down the IRB approval process dramatically, and affect IRB receptivity and the timely approval of future proposals. My purpose in this paper is to discuss these issues, in search of ways for reducing the inherent conflict now apparent between IRB's and surveyors. This paper includes real-world examples and suggests that a solution to this significant conflict relies on finding a balance between providing adequate protection for potential respondents, while protecting survey designs that will produce valid results.

A Qualitative and Quantitative Exploration of the Use of Text Messaging and Emailing on Perceptions of Privacy and Confidentiality

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

Aleia Clark Fobia, *U.S. Census Bureau*

Ryan King, *U.S. Census Bureau*

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

The U.S. Census Bureau conducted the 2014 Census Test during summer 2014 in the Washington, D.C. metropolitan area to test new methods for implementing and conducting the 2020 census, such as connecting physical addresses with third party email and phone number data. This data could be used to contact households in alternative ways, such as emailing or text messaging, that might encourage self-response to the census. However, little is known about how the public would react to receiving an email or a text message from the Census Bureau. Skepticism about the authenticity of a contact or a perception that the government is invading personal privacy could discourage response if not properly addressed. We conducted both quantitative and qualitative research to learn about the impact of these contact methods on perceptions of privacy and confidentiality. As a follow-up to the 2014 Census Test, we conducted seven focus groups with members of households that were sampled for the test. The groups were organized by response type into three rounds, with one round of self-responders, one round of non-responders, and one round of late responders. The focus groups provided an in-depth look into participant perceptions and expectations of privacy and confidentiality during the 2014 Test, and more generally. They also suggested ways in which these concerns might be mitigated. While privacy and confidentiality concerns varied by 2014 Census Test response type, broad concerns about the legitimacy of government emails and dislike of text messaging were shared across response type. Quantitative results from a nightly poll the Census Bureau conducted with Gallup generally support the focus group findings on emailing and text messaging while providing a sense of magnitude and insight into demographic differences in contact strategy perceptions. This paper will present in-depth qualitative focus group findings and quantitative findings from the Census-Gallup poll.

The Value of Personal Information and Privacy: Evidence from Discrete Choice Surveys

Sarah Butler, *NERA Economic Consulting*

Garrett Glasgow, *NERA Economic Consulting*

Recent years have seen rapid growth in privacy-related concerns. While the most high-profile news stories have involved data security breaches at large retailers, there is also a growing awareness of data collection and sharing done by legitimate companies—that is, companies that collect user data and sell this information on to third party advertisers and vendors. Despite the growth of lawsuits involving data collection and data privacy, many areas of the law, public opinion and economics surrounding these matters remain undetermined. One of the central issues in a data privacy case is the potential harm caused by the breach of privacy. But, unless the private or personal data have been used in an unauthorized transaction, the framework for establishing legal and economic harm in this area is unclear. While there are academic papers evaluating the extent to which consumers will allow or prohibit uses of their personal information, empirical data that quantify the value of personal data are rare. There is even less information on how much consumers would be willing to pay to keep their data private or how much they would want to be compensated for the use of their information. Our study fills this gap. In this paper, we use survey and modeling-based techniques to quantify the value that

AAPOR 70th Annual Conference

consumers place on their personal information. To estimate the value that consumers place on their personal information, we designed an experiment using an established survey method called conjoint analysis, which involves giving survey respondents hypothetical choices from which to choose their preferred option. Our results show that in general consumers value their information but there variability in how much consumers care about data sharing. In addition, our results show that a majority of consumers are unaware that many types of data are collected and potentially shared.

Measuring Sensitivity

Robin Kaplan, *Bureau of Labor Statistics*

Erica Yu, *Bureau of Labor Statistics*

A large literature has shown that responses to survey questions can be influenced by sensitivity, or how personal, invasive, or uneasy a question makes respondents feel. Far less is known about the impact of sensitive questions on interviewers. Nonetheless, potentially sensitive questions, for either the respondent and/or the interviewers, may lead to distortions in responses that create threats to data quality. The primary objective of this research is to determine how the perceived sensitivity of survey questions varies across both respondents and interviewers as a first step toward developing a standardized measure of question sensitivity. 780 participants from Amazon Mechanical Turk were recruited to complete an online task. Participants read excerpts from fictional interviews. Questions were taken from a range of federal surveys and selected for their diversity and potential for causing respondents and interviewers to view these questions as sensitive. Topics included employment, alcohol and charitable expenditures, and sleep habits. Within the vignettes, we manipulated two factors between participants: type of sensitivity intervention (positive loading at the start of the question to reduce sensitivity in the form of 'forgiving' wording; negative loading to increase sensitivity in the form of 'unforgiving' wording; or direct wording to have a neutral effect) and the focal vignette character (the respondent or interviewer). Participants then rated the sensitivity of the survey questions using several different sensitivity measures (e.g., "how personal was this question?"). Participants also completed ratings likely to interact with sensitivity, including their attitudes towards the survey topics and how much they empathized with the vignette characters. We will explore the impact of these manipulations to see if they systematically impact participants' judgments of sensitivity across the different measures. We will discuss the implications of our results for measuring sensitivity across respondents and interviewers and working towards a standardized method of measuring sensitivity.

Reassessing and Communicating Concepts of Value and Risk Associated with Survey Data Quality

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

Customary approaches to survey data quality (e.g., Andersen et al., 1979; Groves, 1989; Brackstone, 1999; Biemer, 2010, 2014; Lyberg, 2012; National Research Council, 2013; and Kenett and Shmueli, 2014) often emphasize the multidimensional nature of data quality and note that survey stakeholders expect published statistical series to satisfy quality standards across a wide range of subpopulations and estimands, and across a similarly wide range of population conditions. Within an environment dominated by probability sample surveys with relatively high response rates, the abovementioned norms and standards have led to the development of much of the body of current survey practice. However, as noted by numerous recent authors, continuation of current practice may not satisfy the abovementioned stakeholder needs. This paper suggests as we develop practical methods to address these challenges, we

AAPOR 70th Annual Conference

should do so in a way that accounts directly for the linkage between traditional concepts of survey data quality and concrete measures of value and risk for our primary stakeholders. First, we need to reassess and communicate the linkage between measures of data quality and the practical value that a published statistical series conveys as a form of “public goods” with substantial “option value,” as defined in the traditional economic literature. Some simple survey-related examples include the ability to provide good estimates for a wide variety of survey variables and subpopulations, per Brick (2011); and the ability to produce reliable estimates for parameters that were not anticipated during the design stage, per Fuller (1999). Second, we need to articulate a more refined characterization of the costs and risks associated with failures in one or more dimensions of our survey quality measures. In particular, we would benefit from adaptation of ideas developed originally in the literature on risk management and on “recoverable” or “fault tolerant” designs in engineering and computer science.

Disclosure Avoidance Techniques at the U.S. Census Bureau: Current Practices and Research

Amy Lauger, *U.S. Census Bureau*

Billy Wisniewski, *U.S. Census Bureau*

Laura McKenna, *U.S. Census Bureau*

The U.S. Census Bureau collects its survey and census data under Title 13 of the U.S. Code, which promises confidentiality to its respondents. The agency also has the responsibility of releasing data for the purpose of statistical analysis. In common with most national statistical institutes, our goal is to release as much high quality data as possible without violating the pledge of confidentiality. We apply disclosure avoidance techniques prior to releasing our data products publicly to protect the confidentiality of our respondents and their data. This paper discusses the various types of data we release, our disclosure review process, restricted access procedures, current disclosure avoidance techniques, and current disclosure avoidance research, including the expanded use of synthetic data and remote access systems to increase the amount and quality of data we are able to release.

Monitoring and Evaluating Survey Quality

Establishment of a Quality Management Program: Strategies and Challenges to Implementation

Robyn Sirkis, *USDA National Agricultural Statistics Service*

Pamela McGovern, *USDA National Agricultural Statistics Service*

The National Agricultural Statistics Service (NASS) is committed to providing timely, accurate, and useful statistics in service to U.S. agriculture. An effective quality management program aids in ensuring the utility, objectivity, and integrity of the statistical information that an agency provides to its customers and stakeholders. Therefore, it is essential to establish a quality management program that achieves consistent results and builds quality into survey and census processes. To meet its quality objectives, NASS developed a quality management model for the agency. The key model components include establishing a quality assurance framework of standards and guidelines, implementing and strengthening quality assurance practices, conducting quality assessments and obtaining customer feedback, measuring and improving processes through continual improvement, promoting education and outreach, and ensuring a sense of quality commitment. This paper discusses the quality management model

AAPOR 70th Annual Conference

components and provides examples of the work NASS is doing to ensure product and process quality. We will describe several current and future NASS projects and how each fits into the quality management program, as well as, the opportunities and challenges to implementation. These efforts involve establishing a comprehensive set of statistical standards to instill quality in our information products and promote transparency to our customers and stakeholders; conducting self-assessments of survey processes to establish baseline measures and identify process improvement opportunities; evaluating call history files and setting call parameters to balance respondent burden, cost, and quality; and developing innovative and more efficient sampling and data collection strategies for implementing quality control programs. Finally, NASS is also researching better ways to use data to make informed decisions. Using data and quality metrics to manage and improve processes and products will help NASS achieve greater efficiencies and improved data quality.

A New Computational Tool to Detect Random Responses in Surveys

Delia Dumitrescu, *University of Gothenburg*

Johan Martinsson, *University of Gothenburg*

Sebastian Lundmark, *University of Gothenburg*

Razvan Gurau, *CPHT, Ecole Polytechnique*

Recent work shows that non-differentiation scores are unable to accurately discriminate between random and non-random responses to survey batteries (Dumitrescu and Gurau, 2014). In this paper we develop a new indicator to improve the detection of random response strategies in surveys. We combine formal theory with empirical analyses. Our formal approach is to test the distribution of observed survey responses against a multinomial distribution which describes survey data resulting from random response strategies. We then propose a technique that can produce a more confident assessment that the observed survey data was not produced by a process of random response. The new indicator is implemented in Stata and R. We test its validity against more established non-differentiation indicators that are generally assumed to indicate data quality. Using data from observational studies, Monte Carlo simulations and experimental studies, we show that the new metric performs better than previous indicators at assessing question response quality.

A Demographic Data Quality Monitoring System at the U.S. Census Bureau

Rachel Bray, *U.S. Census Bureau*

Adriana Hernandez Viver, *U.S. Census Bureau*

The U.S. Census Bureau has developed a system to upload each day's Computer Assisted Personal Interviewing (CAPI) data from interviewer laptops to a set of outputs that allow analysts to monitor data quality in nearly real time. The system consists of a series of graphical outputs such as heat maps and daily data "EKGs". The system was developed through a joint effort between statisticians and programmers and has resulted in an additional tool for analysts to monitor data throughout the phase-in of large design changes. The system was used to monitor data quality during the 2010 Sample Redesign for the Current Population Survey and for the National Crime Victimization Survey throughout its 2014 Sample Boost. This paper gives an overview of the system, and provides some preliminary results from the monitoring of the Current Population Survey and National Crime Victimization Survey.

Alternative Data Quality Indicators and Tools to Guide Adaptive Design

Debra L. Wright, *Mathematica Policy Research*

Amang Sukasih, *Mathematica Policy Research*

Michael Sinclair, *Mathematica Policy Research*

Shilpa Khambhati, *Mathematica Policy Research*

Brendan Kirwan, *Mathematica Policy Research*

The tenants of adaptive survey design (Groves and Heeringa 2006) have gained broad acceptance, including a recent endorsement by the Office of Management and Budget (Harris-Kojetin 2014), and have prompted a survey organizations to better refine their data tracking tools and paradata systems. However, clients remain fixated on monitoring response rate as the primary indicator of data quality during (and after) data collection. This is despite research suggesting that other indicators such as variability in key survey estimates and R-indicators (Schouten, et al. 2009) are more effective in terms of informing data collection strategies. In this paper we present tools we designed to track these alternative measures in an ongoing study, in particular we track comparability between treatment and control groups and sample representativeness during data collection. Examples of data collection reports and a data collection strategy that utilizes such reports to made mid-course corrections to the data collection strategy will be presented.

Explaining Variation in Monitors' Detection of Interviewing Errors in Telephone Surveys

Doug Currivan, *RTI International*

Paul Biemer, *RTI International*

Tamara Terry, *RTI International*

Gordon Brown, *RTI International*

To limit the potential for interviewer behavior to add variance or bias to survey estimates, monitoring telephone surveys requires accurate and consistent detection of interviewing errors, interview protocol violations and other biasing behaviors. Multiple factors could potentially affect the detection of telephone interviewing errors, including monitor characteristics, interviewer characteristics, interviewer supervision, the survey protocol, the call center environment, and the monitoring system. RTI's standardized, mode-independent interview quality monitoring evaluation system (QUEST) allows telephone interviewing behaviors to be evaluated in both live and recorded sessions using a common set of quality metrics based on empirically derived indicators. The standardized, empirical focus of QUEST is intended to promote accuracy and consistency in the detection of interviewing errors. Despite this goal, variation in monitors' detection of interviewing errors is typically observed and the source of this variation is not fully understood. Building on the authors' prior work, this study examines to what extent the observed variation in interviewing errors across QUEST skill areas can be attributed to monitors' behavior alone. Using multi-level models, the analysis also assesses the additional impact of other factors – monitor characteristics, interviewer characteristics, and live versus recorded sessions – on variation in detected interviewing errors. The analysis combines data from (1) QUEST monitoring sessions, (2) administrative records with monitor characteristics, and (3) administrative records with interviewer characteristics for three telephone surveys. These data allow for operationalization of multiple sources of variation in monitors' error detection that cannot be captured in the monitoring results alone. The multi-level models analyze the contribution of monitors' behavior, monitors' characteristics, and interviewer characteristics to variation in the detection of interviewing errors. This presentation discusses the implications of

AAPOR 70th Annual Conference

these results for understanding sources of variation in monitors' detection of interviewing errors and for guiding decisions on designing monitoring processes in centralized telephone survey centers.

Panel: Public Opinion and Energy Policy: Pipelines, Nuclear Power and Fracking

Panel: Public Opinion and Energy Policy: Pipelines, Nuclear Power and Fracking

Erick Lachapelle, *Université de Montréal*

Timothy B. Gravelle, *Univeristy of Essex*

Christopher Borick, *Muhlenberg College*

Issues such as energy self-sufficiency, energy conservation and the building of new energy infrastructure rate high on the current policy agenda, not only in the United States, but also in other national contexts. In particular, creating and transporting the energy resources needed to meet public demand has become highly politicized. In such a context, it is essential to understand the dynamics of public opinion toward energy policy issues. The panel focuses current research on the construction of contentious new oil pipelines (such as Keystone XL), nuclear power, and hydraulic fracturing. Presentations draw on a range of data sources: nationally representative surveys, state surveys, and media content analyses. The panel brings together public opinion researchers from academia and the private sector; it also brings together American and Canadian researchers to discuss public attitudes toward energy issues in cross-national perspective.

Framing and the Pipeline: Analyzing the Persuasiveness of Arguments for and Against the Keystone XL Energy Pipeline in Canada and the United States

Erick Lachapelle, *Université de Montréal*

Timothy B. Gravelle, *Univeristy of Essex*

Christopher Borick, *Muhlenberg College*

The North American continent is covered in an extensive network of energy pipelines that crisscross inter-provincial, inter-state and international borders. While largely invisible to the public, proposals for the building of new energy pipelines have become a flashpoint for political controversy in recent years. An example is Trans Canada's infamous Keystone XL energy pipeline proposal that has divided Canadians and Americans alike. Despite having very different interests in getting the pipeline built, the political rhetoric in these two highly interdependent, geographically contiguous countries is framed largely as a debate between jobs and the economy vs. local risks and global climate change. In this context, this paper analyzes results from a survey framing experiment designed to test the persuasiveness of different arguments made for and against the Keystone XL energy pipeline, administered to two nationally representative samples of the American and Canadian populations in the Fall of 2014. Results build on the framing literature, and shed light on which arguments are more persuasive for mobilizing different groups for and against energy development projects in two entirely different economic-environmental contexts.

Energy Knowledge as a Predictor of Attitudes toward Nuclear Power

Joe Murphy, *RTI International*

Brian Southwell, *RTI International*

Sarah Parvanta, *RTI International*

Public opinion on nuclear power is an important consideration for energy policy and local adoption as we consider alternatives to fossil fuels. Past research has tracked attitudes towards nuclear power over time, e.g., showing an increase in support for increased use of nuclear power in the U.S. from 28% to 34% between 2002 and 2007, and investigated factors associated with support of nuclear power, including cost, local environmental risks, likelihood of an accident, and concerns about waste storage and climate change. It appears attitudes can shift in the short term in reaction to events such as the Fukushima Daiichi nuclear disaster. What has not been considered previously is the role of energy knowledge with regard to opinion on nuclear energy more generally. That is, are those who are more knowledgeable about energy more or less likely to view nuclear power favorably? We explore this question using data from the 2011 Energy Behavior, Knowledge, and Opinions Survey, fielded to a nationally representative sample of 816 U.S. householders. We measured both favorability of nuclear power and knowledge via an 11-item energy knowledge index. Findings indicate significantly higher rates of formed attitudes and support for nuclear energy production as an alternative to coal or natural gas among those with high levels of energy knowledge compared to their peers. This suggests a potentially important role for energy education and literacy in explaining gaps between groups in the formation of nuclear energy attitudes.

Same Old NIMBY Phenomenon? The Effect of Proximity on Americans' Perceptions about Hydrofracking

Matthew Barnes, *West Virginia University*

As the shale oil and natural gas boom continues across the United States, controversy follows in its wake. Opponents have called for prohibitions on hydraulic fracturing, and many local governments have instituted bans or moratoria on the practice, raising questions of whether opposition to hydrofracking is simply a NIMBY ("not in my backyard") response. Current research on public attitudes toward hydrofracking is limited, and existing studies do not adequately measure respondents' geographic proximity to drilling activity. Linking the location of hydraulically fractured wells to survey respondent geographic location, I examine whether individual opposition toward hydrofracking is determined most strongly by proximity to drilling activity or if other factors like socio-demographics, partisanship, or environmentalism better explain attitudes. I conclude by discussing the policy implications of individual attitudes toward fracking to local, state, and national energy policies.

Developing and Testing a Framework for Understanding Public Support of "Fracking"

Jessica Alcorn, *Indiana University*
Olga Schenk, *Indiana University*
John D. Graham, *Indiana University*
John Rupp, *Indiana University*
Sanya Carley, *Indiana University*
Michelle Lee, *Indiana University*
Yu Zhang, *Indiana University*
Ashley Clark, *Indiana University*

As citizens within many U.S. states have interacted with increased natural gas production through unconventional gas development (UGD), the public's attitudes towards the practice have not been straightforward. Previous research on acceptance of energy technologies, and UGD specifically, highlight a wide range of factors that may drive public attitudes, namely sociodemographic characteristics and political ideology. Decades of research have also sought to explore the dimensionality of environmental attitudes; however, the theoretical frameworks and methods from this literature have not been employed in analyzing attitudes towards specific kinds of energy development. In this study, we conducted a survey of adult U.S. residents in six U.S. states where UGD is underway or geologically promising: three states with high and/or growing production (Ohio, Pennsylvania, and Texas) and three that have little or no UGD (New York, Illinois, and California). We analyze both whether respondents typically identify different components of advantages and disadvantages of UGD in their attitudes and what factors influence attitudes. Our analysis comports with prior research indicating that those who identify as White or male perceive fewer disadvantages of UGD. As expected, older respondents and those who receive royalty payments perceive more advantages of UGD, and those who identify as members of the Democratic Party perceive more disadvantages of UGD. However, the role of education in influencing attitudes remains somewhat unclear. With respect to dimensionality, we find that respondents generally do not differentiate between the various aspects of potential advantages and disadvantages of UGD.

Methodological Briefs: I'll Do Anything for a Better Response Rate

Using the Study Sponsor Name in CATI Introductions: Effects on Call Outcomes and Eligibility Rates

Mehera Baugher, *Abt SRBI*
Martina Smith, *Abt SRBI*
Kelly Daley, *Abt SRBI*

In RDD surveys, the introductory script is typically the first point of live interaction with a sampled household. Engaging the household respondent at the initial contact is critical for successful screening and completion of the survey. Among many factors that influence a respondent's decision to cooperate, perception of the study sponsor's authority is critical: if the study sponsor is not perceived as having legitimate authority, the respondent may be less willing to cooperate (Groves, Cialdini, and Couper 1992). The 2014 Worker Classification Survey is a large, dual-frame RDD survey of workers age 18 and over sponsored by the U.S. Department of Labor and conducted by Abt Associates and Abt SRBI. The survey measures workers' knowledge about their current job classification (employee, independent contractor or

AAPOR 70th Annual Conference

something else), and their knowledge about the rights and benefits associated with their job status. We conducted an experiment on over 21,000 cases to test the effect of sponsor name in the study introduction. Specifically, we randomly assigned half of the cases in the experiment to identify the study contractor (Abt Associates) in the introductory sentence and the other half to identify Department of Labor. In this brief, we present results of the experiment. We hypothesized that conspicuous reference to a regulatory agency might lead to self-exclusion (screen-outs). Conversely, an introduction that includes the name of the government agency would be more widely recognized and perceived as a legitimate authority, consequently yielding higher cooperation rates. Our analyses compare call outcomes and eligibility rates between the two introductory conditions. Results for some call dispositions were significant, suggesting that using the name of the government agency, rather than the private company, lead to higher cooperation and completion rates, but also to lower rates of eligibility. We discuss the implications of these results for other RDD studies.

Does a Pre-Notice Letter Affect Response in the Internet Collection Mode in the American Community Survey?

Padraic A. Murphy, *U.S. Census Bureau*

Andrew W. Roberts, *U.S. Census Bureau*

The Census Bureau's American Community Survey (ACS) added an Internet data collection mode as part of its sequential design in the January 2013 data-collection cycle. Previously, the initial mode of collection was a mail-back paper questionnaire, and a few days before the first mailing of a questionnaire, the Census Bureau would send a pre-notice letter. It is well established in the survey literature that such a pre-notice letter for mail-back surveys significantly improves response. The ACS continued the use of a pre-notice letter with the Internet mode; however, the 2012 National Census Test included research on Internet data collection, and found no significant difference in response with or without a pre-notice letter. Therefore, in 2014, the ACS conducted a test to see if dropping the pre-notice letter would have any effect on Internet response and total self-response (Internet and mail-back together). We will discuss the results of the test, comparing the Internet response rate for an experimental group where the pre-notice letter was dropped with the rate for a control group where it was kept. We will also compare the total self-response rates for the two groups. Finally, we will discuss a cost analysis that compares the average total cost per case between the two groups.

The Effects of Persuasion Messages in the Advance Letter

Cong Ye, *American Institutes For Research*

An analysis of response rates over time found that both the noncontact rate and refusal rate had been increasing during the 20-year analysis period, but the refusal rate increased at a greater rate (de Leeuw and de Heer 2002). The efforts to gain cooperation from the sampled member are crucial in achieving or maintain a high response rate. One method is to send persuasion letters to the sampled members. This study examined the effects of priming persuasion techniques in the advance letter (using a subtle message as the stimulus to influence the sampled member's response to the survey) on response rates to a mixed-mode panel survey which included a sample of originally 1,500 households. In the advance letters, one quarter of the sample members received an additional sentence "your responses in previous survey show that you are a helpful person" (helpful priming), another quarter received an additional sentence "almost everyone like you responded in the last wave of the survey" (norm priming), another received both, and the final quarter was a control group, receiving neither priming message. The analysis compared the differences in response rates, demographics, and responses among the

AAPOR 70th Annual Conference

experimental groups. The results show that the persuasion messages need to avoid obvious priming as it may work against the efforts.

Formal or Friendly: Does Messaging Style Impact Survey Response?

Kerry Y. Levin, *Westat*

Jocelyn Newsome, *Westat*

Stephanie Beauvais Dennig, *Westat*

Brenda Schafer, *Internal Revenue Service*

Pat Langetieg, *Internal Revenue Service*

Melissa Vigil, *Internal Revenue Service*

Michael Sebastiani, *Internal Revenue Service*

Much like other government surveys (Groves 2011), the IRS Individual Taxpayer Burden (ITB) Survey has generally faced declining response rates over recent years. The ITB Survey is a nationwide survey conducted annually with about 20,000 taxpayers. In particular, response rate to the ITB Survey has been low among Millennials, who historically respond at lower rates (Molyneux & Teixeira 2010). There are many different ways researchers have attempted to address declining response rates. Dillman (2014) has discussed the importance of the overall communication strategy when administering a survey, including careful consideration of the visual design of all materials. He advocates “respondent-centric” design that appeals to a wide variety of people. With this in mind, we conducted an experiment to examine the impact of different messaging styles on response rates. We compare a more formal style with one designed to appear more open and engaging. This friendlier messaging also included an appeal to civic responsibility and improving the status quo, which has been shown to be particularly effective for Millennials (Molyneux & Teixeira, 2010). In our experiment, half of the respondents received the formal approach, which emphasized “official” insignia such as the IRS logo. The other half received a less formal, more casual approach using graphics and phrases that emphasized “Your experience counts” and “Take the survey, make a difference.” The IRS logo was still prominently featured to assure respondents that the survey was legitimate. This paper will explore the relative success of each strategy by looking at overall response rates, as well as the impact on specific populations of interest.

Maximum Bang for Minimum Buck: Using Complex Incentive Targeting Strategies to Improve Panel Compliance

Darin D. Harm, *The Nielsen Company*

Christine Heiss, *The Nielsen Company*

Previous research has shown that offering monetary incentives to respondents can increase desired respondent behavior (e.g., survey response, behavioral compliance, etc.). There is less research on the effects of short-term incentive increases in longitudinal panels. This study examines the effects of a short-term monetary incentive on panelist behavioral compliance. Nielsen uses a multi-stage mixed-mode approach (mail, phone and in person) to recruit households into a two-year panel. People are asked to wear a small, cell-phone sized device that passively collects exposure to media throughout the day. Encouraging people to wear the device throughout the day is critical so we can include the information each day. In June 2014, panelists in selected households were offered a targeted monetary incentive for consistently wearing their device during a two-week period when compliance is typically lower than average. Compliance decreases during this summer time period for a variety of reasons (e.g., forgetting to wear the device while on vacation, not wanting to bring the device to the beach, etc.)

AAPOR 70th Annual Conference

Panelists were selected for the targeted incentive based on a variety of indicators, including presence of children, time in the panel and historical performance data about how often they wear the device. To quantify the impact of this targeted incentive, a random sample of eligible households was excluded from the targeted incentive to serve as the control group. The results of this test will be discussed, along with the implications of using targeted incentives (monetary and non-monetary) to cost-effectively improve the quality of data. The benefits and limitations of this approach to improve compliance with particular behaviors will be reviewed, along with opportunities for future research.

Can a Magnet Attract Respondents? Incentive Type and Monetary Value Effects in an RDD Survey

Matt Jans, *UCLA Center for Health Policy Research*

John Rauch, *Westat*

Sherman Edwards, *Westat*

David Grant, *UCLA Center for Health Policy Research*

Royce Park, *UCLA Center for Health Policy Research*

Pre-paid cash incentives increase survey response rates in every mode, yet we still know little about the social and psychological mechanisms at work beyond their non-economic nature. Non-monetary incentives may activate social psychological mechanisms thought to be responsible for response (e.g., social exchange, norm of reciprocity, liking, social benefit, and positive affect), but cash incentives usually out-perform cash equivalents and gifts (Cantor, O'Hare, and O'Connor, 2008). Yet, recent evidence shows that a refrigerator magnet with the sponsor's logo can perform as well as a \$5 pre-paid incentive in a mail survey (Brick et al, 2012). To explore this novel finding further, we conducted a pseudo-replication in the California Health Interview Survey (CHIS). CHIS is a RDD telephone survey that mails advance letters to address-matched landline sample units prior to calling. We randomized address-matched landline sample units to four pre-notification incentive groups: \$2 cash, \$5 cash, thin magnet, or thick magnet. The magnets were about business card size and displayed the CHIS name and logo. Varying magnet thickness and weight allowed us to test whether the feel of the envelope increased cooperation by making the letter more likely to be opened. Results do not replicate Brick et al's findings, but support expectations about cash v. non-cash incentives. Cash incentives had initial cooperation rates more than 8 percentage points higher than the magnets. Thin and thick magnets performed about the same. The \$2 condition was 1.4 percentage points lower in cooperation than the \$5 condition (non-significant), which conflicts with past research showing that \$5 out-performs \$2. Differences between the Brick et al study and this experiment may be responsible for the failure to replicate including features of the magnet and response task. Proposals for future research exploring incentive effect mechanisms will be discussed.

Investigating the Relationship Between Nonmonetary Incentives, Questionnaire Length and Response Rates in a Physician Survey

Eric Jamoom, *National Center for Health Statistics*

Paul Beatty, *U.S. Census Bureau*

This paper will present new findings from an experimental program on the effect of nonmonetary incentives in mail surveys of physicians. Previous experiments from this program (presented at the 2013 and 2014 AAPOR Conferences) showed that including high quality pens with initial mailings boosted physician response rates in several surveys. The current experiment explores whether the effect of nonmonetary incentives varies based on length of self-administered

AAPOR 70th Annual Conference

questionnaires, and whether nonmonetary incentives offset reductions in response rates due to questionnaire length. Data come from the 2014 National Electronic Health Record Survey, a mail survey with telephone follow-up of office-based physicians, which is used to measure physician adoption and use of electronic health records. In 2014, two versions of the questionnaire were developed: a 4 page questionnaire, and an 8-page questionnaire that included all of the questions on the 4-page version, plus additional content. We randomly assigned 10,302 physicians into 4 treatments: 4-page questionnaire with a pen, 4-page questionnaire with no pen, 8-page questionnaire with a pen, and 8 page questionnaire with no pen. In addition to distinguishing between the effects of questionnaire length and the nonmonetary incentive on final response rates, study findings will demonstrate how response rates change over the course of the survey cycle, from the initial mailing (which included the pen for respondents in those conditions) through follow-up mailings, and finally through attempts to contact initial non-respondents via telephone.

A Response Rate That Was Too Good to be True: Detecting and Explaining Fraud in a Pharmacy Based Consumer Web Survey

Peter Batra, *College of Pharmacy University of Michigan*

Attitudes and behaviors toward a pharmacy-based intervention of the prevention of unplanned pregnancy were assessed through a web-based survey. Data collection over three time periods captured cross sectional responses prior to the intervention (time 1), during the intervention (time 2) and after the intervention was completed (time 3). The intervention comprised various informational programs aimed at increasing awareness in preventing unplanned pregnancies for consumers visiting the pharmacy. During each of the 3 time periods, pharmacy staff distributed post cards to pharmacy customers containing the web address inviting participation in a 20-question self-administered survey that took on average 5 minutes to complete. An electronic gift card was used as an incentive for completing the survey. Approximately half way through the 2 month data collection window, we detected record level non-unique responses in the dataset. This was investigated mainly as skepticism that a higher than average response rate was mounting. Fraudulent survey completion was discovered as a result of looking at survey metadata such as ISP address information and survey start and end times. This methodological brief will highlight the specific process we used to distribute recruitment materials, detail how the fraudulent activity was discovered and quantified and which records were removed from our dataset. There will also be a summary of the lessons learned from this experience.

Finding the Best Time to Make Contact Attempts for Face-to-Face Interviews

Dato Tsabutashvili, *Gallup*

Neli Esipova, *Gallup*

In face-to-face surveys interviewers' travel expenses makes up a huge part of survey costs. Minimizing the number of visits required to find desired respondents is crucial for optimizing budgets of for face-to-face surveys. To this purpose, it is very useful to know which days of the week and what time of the day are respondents are most likely to be at home and available for interviews. The first purpose of this paper is to identify which day of the week and which what time of the day has the best response rates. We expect that different times might be best for different countries and for different types of settlements within countries. The analysis will be based on Gallup World Poll data of several countries from Eastern Europe and Central Asia where face to face interviews are a widespread survey method. We will use data from contact sheets where interviewers were asked to record the time and date of each contact attempt done during fieldwork. The second purpose of the paper is to examine how the respondents

AAPOR 70th Annual Conference

contacted and interviewed during different times of the day differ in terms of key demographics. We expect that choosing visit times carefully can improve the representativeness of the sample. Using the key demographics such as employment, gender, age, household size, and education from official population censuses as targets to compare, we will examine which times of the day are most likely to produce the best representative samples. The analysis will be done on at the country level and at the settlement type level (big cities, smaller towns, and rural settlements).

Saturday, May 16
8:00 a.m. – 9:30 a.m.
Concurrent Session G

Mini-Conference: Combining Probability and Non-Probability Samples

Combining a Probability Based Telephone Sample with an Opt-in Web Panel

Randal ZuWallack, *ICF International*

James Dayton, *ICF International*

Naomi Freedner-Maguire, *ICF International*

Katherine J. Karriker-Jaffe, *Alcohol Research Group*

Thomas K. Greenfield, *Alcohol Research Group*

Our research assesses the feasibility of using data fusion to combine data for an alcohol survey collected from a dual-frame random digit dialing (RDD) telephone sample, with data collected from an opt-in web panel. A benefit of RDD telephone samples is that they are based on probability sampling, yet they are challenged by decreasing response rates and increasing costs. Data collection via opt-in web-panel raises concerns regarding population representativeness, yet is a fraction of the cost of a telephone sample. A hybrid approach that fuses RDD survey data with opt-in panel data is attractive from a cost perspective. The source of the RDD data is the thirteenth iteration of the National Alcohol Survey (N13), a dual-frame telephone survey conducted on landline and cell phones. N13 covers a number of alcohol related topics, including alcohol consumption and behaviors, effects of alcohol on individual lives and the lives of others, expenditures for alcohol, alcohol-attributed and non-attributed health conditions and perceptions about alcohol, as well as related factors such as emotional well-being. We hypothesize that a hybrid approach will offer probability-based prevalence estimates for drinking behaviors (e.g. drank wine in past month) and health conditions, yet allow us to collect episodic data such as volume consumed and brands, as well as expenditures for alcohol via an abbreviated version of the N13 questionnaire administered to respondents from a web panel. Our presentation focuses on identify the linking variables to fuse the data and assess conditional independence, a critical assumption for data fusion. We then compare responses from the web data with the telephone data, controlling for the linking variables. These comparisons will increase our understanding of whether web panel responses will produce estimates comparable to the telephone given that there could be mode effects and/or differences due to population representativeness.

AAPOR 70th Annual Conference

Fit for Purpose: Supplementing RDD Surveys with Online Opt-In Panels

Robert P. Agans, *University of North Carolina at Chapel Hill*

Increased costs and nonresponse have lead population-based investigators to explore non-probability sampling methods. The 2011 AAPOR Task Force on Non-probability Sampling has recommended that such strategies be evaluated on fit for purpose or such characteristics as relevance, accuracy, timeliness, accessibility, interpretability, and consistency. The present study will evaluate each of these criterion by comparing sample gathered via a national dual frame RDD survey (n=3,800), with over half of the sample coming from a RDD cell phone frame, and an Amazon Mechanical Turk (MTurk) web-based opt-in panel (n=2,742). Emphasis will also be placed on developing proper statistical techniques to combine both probability-based and nonprobability-based approaches into one robust sample for estimation purposes. Implications for oversampling hard-to-reach populations and potential costs benefits will be explored as well.

A Comparison of Online Panels with GSS and ANES Data

Elizabeth S. Zack, *Department of Sociology, Indiana University Bloomington*

John M. Kennedy, *Center for Survey Research, Indiana University*

In the past five years, researchers have increasingly used low cost data collection methods to conduct surveys. Survey data collection software platforms such as Qualtrics and Survey Monkey allow researchers to easily and cheaply create questionnaires for distribution. Similarly, low cost methods are available to recruit survey participants. Some of these include online panels, Amazon's Mechanical Turk (MTurk), and Google Consumer Surveys. With these tools, researchers are much less dependent on professional survey researchers to conduct surveys. More importantly however, survey researchers are now using non-probability online panels as a substitute for probability samples. In 2010 and 2013, AAPOR released task force reports that analyzed the challenges encountered when using online panels and nonprobability samples for high quality survey research. In 2014, a book on the use of online panels in survey research included chapters written by respected survey researchers (Callegaro et al, 2014). In the past five years, at least 20 peer-reviewed methods articles were published on the use of Mechanical Turk for social and behavioral science research. Despite the cautions raised about the appropriate use of online panels, they are being used more often, e.g., the YouGov panel and CBS News. In this presentation, we will discuss the results of a number of experiments we conducted that compared distributions in questions asked recently in the ANES and GSS to similar questions asked with MTurk samples and a Qualtrics online panel. In addition, we will show how simple multivariate models are similar and different using data from both the probability and non-probability samples. This presentation will contribute to the continuing research into the appropriate uses of online panels for survey research.

Can a Non-Probability Sample Be Used to Measure Emerging Tobacco Product Use Among Young Adults?

James J. Dayton, *ICF International*

Tala Fakhouri, *ICF International*

Despite the decline in the prevalence of tobacco use in recent years, as many as 25% of adults still smoke. Additionally, the use of emerging tobacco products such as hookah, snus, and electronic nicotine delivery systems (ex. e-cigarettes) is gaining popularity among U.S. young adults. These products are thought to promote the initiation of nicotine addiction and may lead to combusted tobacco use among non-users. Currently, there is limited information on the diversity of tobacco products used by young adults, and the impact of tobacco product

AAPOR 70th Annual Conference

marketing on use behaviors, perceptions, and attitudes. In collaboration with Emory University and the University of Texas at Austin, ICF is operating two longitudinal surveys to better understand the epidemiology of emerging tobacco use and the effect of marketing on use among U.S. young adults enrolled in universities and colleges in Georgia and Texas, respectively. Ideally, a sample of all young adults in Georgia and Texas would have been recruited; however, a cost effective probability-based sample was not available. ICF is testing a lower cost sampling alternative that may provide a more complete coverage of the young adult population. We use population web panels drawn as Census balanced samples to test the utility of non-probability surveys in both states. ACS and state NHIS data will be used to assess the accuracy of non-probability sampling. A comparison of estimates on emerging tobacco product use and exposure to marketing will be conducted between the non-probability sample, and the school-based sample.

Evaluating a Propensity Score Adjustment for Combining Probability and Non-Probability Samples in a National Survey

Kurt R. Peters, *ICF International*

Heather Driscoll, *ICF International*

Pedro Saavedra, *ICF International*

Research designs that make use of non-probability samples offer the possibility of fit-for-purpose data at a fraction of the cost of probability-based designs. In 2012, ICF designed and fielded a national survey in Canada that gathered data about Canadians' 2011 nature-based recreation, subsistence, and conservation activities. To support larger sample sizes at the provincial level, a hybrid design was employed that combined a stratified national probability sample of more than 76,000 adults randomly selected from an addressed-based sample (ABS) with a non-probability sample of adults randomly recruited from a purchased Web panel. As part of the survey weighting, ICF developed a propensity score adjustment to combine the data from the probability and non-probability samples. This presentation will evaluate the effectiveness of the propensity score adjustment in decreasing bias in key demographic and survey outcomes in the non-probability sample. We will also discuss implications concerning the effectiveness of hybrid designs and the degree to which non-probability data could be considered fit-for-purpose in similar survey applications.

Incentive Effects

Non-Monetary Incentives Related to the Survey Topic and Survey Participation

Nicholas Ruther, *Abt SRBI*

Dianne Rucinski, *Abt SRBI*

David P. Eisenman, *Assc Prof of Med. and Pub. Hlth UCLA, Preparedness Science
Officer LA Cty Dept of Pub. Hlth*

Non-contingent monetary incentives have demonstrated their effectiveness in increasing response rates (Singer & Ye 2013; Cantor, O'Hare, & O'Connor 2008), and potentially among respondents for whom the survey topic has low salience (Groves et al 2006; Groves, Presser, & Dipko 2004; Groves, Singer, & Corning 2000). However, not all funding agencies and investigators are amenable to the use of non-contingent or monetary incentives. To maximize the effectiveness of contingent non-monetary incentives, some researchers select items related to the survey topic, such as diapers for new parents or school supplies for students. The

AAPOR 70th Annual Conference

potential weakness of this approach is that tailoring the incentives to the survey topic may differentially increase the response rate among those most interested in the survey topic rather than encourage participation among less interested respondents. Our research is based on a telephone survey of disaster preparedness in 16 communities in Los Angeles County to examine the effectiveness of contingent nonmonetary incentives related to the survey topic. Whereas all respondents were offered their choice of an emergency supply item, a portion of respondents who completed the survey declined the offer. This study uses non-incentive respondents as a proxy for non-responders under the hypothesis that both groups are likely to share lower interest in the incentives. Preliminary analyses indicate that accepting the incentive was associated with interest in the topic. Respondents who received incentives were more likely to believe planning for disasters matters (86% vs. 75%), planning with neighbors will help (78% vs. 66%), and planning with neighbors is necessary (92% vs. 87%). They were also more likely to report looking for information on preparing (36% vs. 30%). This investigation sheds light on a novel aspect of survey incentive and topic interaction.

The Effect of Large Monetary Incentives on Survey Completion Costs: Evidence from a Randomized Experiment

Joanne Hsu, *Federal Reserve Board*

Maximilian Schmeiser, *Federal Reserve Board*

Cathy Haggerty, *NORC*

Shannon Nelson, *NORC*

The Federal Reserve Board's triennial Survey of Consumer Finances (SCF) collects information about family incomes, net worth, and other financial outcomes. As with other national surveys, over time it has become increasingly difficult to reach the SCF's target response rate. In 2014, the Board partnered with NORC to conduct a field experiment to test how monetary incentive payments to respondents affects response rates and the amount of effort required to obtain SCF interviews. The experiment involved conducting a survey that was billed as the 2014 SCF. An area probability sample of 900 randomly selected residential addresses was drawn from three large U.S. cities. Half of all respondents received \$5 as a prepaid incentive with their invitation letter, while the other half received invitations without prepaid incentives. Sample addresses were randomly assigned to one of three post-paid monetary incentive amounts for completing the SCF: \$50, \$100, and \$150. Finally, half of the respondents in each category were assigned to receive an escalated offer on their second contact from an interviewer should they decline to respond to the SCF on their first live contact with an interviewer: \$75, \$150, and \$250. Prospective respondents received the exact same information and recruitment strategies as previous SCF waves. Interviewers were blind to the extent possible about the assignment of experimental groups. Preliminary results suggest that the prepaid incentive and higher incentives for completion increase response rates and reduce interviewer effort required to obtain completions, particularly for the \$150 initial incentive. Moreover, escalations in the two higher incentive groups appear to increase response rates at the second contact. Our results provide new information on the effectiveness of large prepaid and postpaid monetary incentives in surveys. This information can be used to determine the optimal balance between monetary incentives and interviewer effort for minimizing survey costs.

The Effectiveness of Using Prepaid Incentives in a Mixed-Mode Survey

Linda S. Mendenko, *Mathematica Policy Research*

Heinrich Hoch, *Mathematica Policy Research*

Priyanka Anand, *Mathematica Policy Research*

Rebecca DiGiuseppe, *Mathematica Policy Research*

Ryan McInerney, *Mathematica Policy Research*

Including prepaid incentives with advance letters requesting survey participation is an increasingly valuable tool to encourage potential respondents from address-based samples. In this context, we investigated whether sending a \$5 bill in an advance letter would affect the response rate for a mixed-mode survey of unemployed workers conducted for the U.S. Department of Labor. The survey featured a web-first survey, with a delayed telephone-response option. Our experiment involved randomly assigning approximately 1,250 sample members to two groups receiving either partial prepayment or total post-payment of incentives. The sample members were from three states and randomized to ensure a balanced distribution of demographic and pre-layoff job characteristics across the two groups. Both groups were offered a total incentive of (1) \$40 for respondent-initiated completions via the web or by phoning in to complete through computer-assisted telephone interviewing (CATI); or (2) \$30 for CATI completions after they were called by the survey team. The prepaid group was mailed \$5 cash in advance of receiving the survey, whereas the other group received the entire payment after survey completion. Our findings include the following: · The \$5 partial prepayment increased the response rate in our target sample by about one-tenth. · This response rate increase is almost entirely due to an increase in web responses. · The size of the prepayment effect peaks about five weeks after a sample release, suggesting an interaction effect between the prepayment and a reminder letter that was mailed in the fourth week of each release. · The prepayment also lowered the share of cases requiring active locating efforts, which would have been costly to conduct. This paper discusses the methodology and results of the prepayment experiment. We present additional analyses of how the prepayment affected costs and provide recommendations for future survey research.

Preliminary Results from an Incentive Experiment for Ohio PRAMS

Marilyn I. Wilkinson, *Abt SRBI, Inc.*

Rebecca Devlin, *Abt SRBI, Inc.*

Melissa VonderBrink, *Ohio Department of Health*

Connie Geidenberger, *Ohio Department of Health*

Katie Labrecque, *Abt SRBI, Inc.*

The goal of the Ohio Pregnancy Risk Assessment Monitoring System (PRAMS) is to improve the health of new mothers and their babies. This multimode mail and CATI surveillance effort has been conducted in Ohio since 1999 with approximately 200 new mothers (1 out of every 62 births) selected on a monthly basis from birth registries. Whereas the Ohio PRAMS typically achieves a weighted response rate of 65%, the response rate among African American mothers is significantly lower at 45% to 50%. The type and dollar value of PRAMS incentives are determined by individual states. Yet, with the exception of two published studies conducted in the Wisconsin and Ohio PRAMS there is little empirical evidence to guide the choice of incentives most likely to increase the response rate overall and among African American mothers in particular. To examine the current impact of prepaid incentive type on the Ohio PRAMS response rate, the prepaid incentive included in the initial mailing for the 2014 Ohio PRAMS was randomly assigned to two conditions: a \$10 Walmart gift card (the existing

AAPOR 70th Annual Conference

protocol) vs. \$10 cash. Contrary to the Wisconsin study which found that prepaid cash incentives were most effective with African American mothers, our preliminary results based on the 600 surveys completed three months into the eight month field period reveal no association between incentive type and response rate comparing Black and White women or different education levels. Rather, among White mothers, those over 30 ($p=.03$), married ($p=.07$), and not on Medicaid ($p=.06$) prefer cash to a gift card based on Chi-square tests. This paper will report the results of the incentive experiment for the entire survey which is scheduled for completion in January, 2015.

Incentive Use Tracking and the Effect of Incentives on Interview Completion for the General Social Survey

Beth Fisher, *NORC at the University of Chicago*

Michael Buha, *NORC at the University of Chicago*

A great deal has been written about the use of incentives in survey research, types and timing of incentives, and how this impacts interview participation rates. Following up on previous work presented at AAPOR's 2012 annual conference, changes were made to incentive tracking in the most recent round of the General Social Survey. Before the start of data collection for the 2014 General Social Survey, modifications were made to case management software to allow for more sophisticated tracking of incentive and 'in kind' gift offers, use, and who these items were offered to. Based on work presented in 2012, we also made a more concerted effort to limit the number of dollar amounts offered to the respondent. Our research questions continue to be prompted by the following anecdote from interviewers: Allowing immediate use of incentives will result in faster completions. From this, we ask: Do non-random incentive payments appear to make a difference in interview completion rates? Does use of in-kind gifts, timing, differentials in the incentive amount, and amounts of incentives offered in previous rounds of the General Social Survey impact response rates for panel and address-based respondents?

Evaluating Polling Accuracy

Evaluation of Mid-Term Election Polling in Georgia

Seth Brohinsky, *Abt SRBI*

Dean Williams, *Abt SRBI*

Courtney Kennedy, *Abt SRBI*

Many polls struggled to accurately predict election results in several 2014 mid-term races. The problem was particularly pronounced in Georgia, where polls failed to accurately gauge the margin of victory in both the U.S. Senate and Gubernatorial elections. Exit poll data indicate that one problem with the Georgia polls was that they under-estimated Republican support among white voters. In an effort to improve future polling in Georgia, we identified four weighting adjustments that may have produced more accurate estimates, had they been implemented. Each of these adjustments stems from the premise that the polls underestimated Republican support among whites – particularly less educated, lower income whites. In this paper, we empirically test the following weighting adjustments: (1) including household income as a post-stratification variable; (2) including party identification as a post-stratification variable; (3) including separate categories for "less than high school" and "high school graduate/GED" as opposed to using a combined category in post-stratification; (4) using historical data (from GA Secretary of State) on the demographic profile of the midterm electorate in the post-

AAPOR 70th Annual Conference

stratification. We analyze whether the various weighting corrections improve the final election estimates in light of the actual result. The data for this analysis come from the Abt SRBI/Atlanta Journal Constitution Georgia poll conducted in late October 2014. As pollsters prepare for the 2016 election, it is critical to understand how the 2014 elections polls can be improved. If the weighting corrections proposed improve the 2014 election estimates, then pollsters should consider implementing these changes for the 2016 election cycle.

Sources of Error in the 2014 Midterm Pre-Election Polls

Natalie Jackson, *Huffington Post/Pollster.com*

Mark Blumenthal, *Huffington Post/Pollster.com*

Polling on key races in the 2014 midterm elections generally forecast the correct winners, but most polls and forecasts overstated support for Democrats and collectively missed the magnitude of the Republican wave. . Many of the races expected to be close turned into Republican blow-outs on Election Day; some that were supposed to be easy Democratic holds turned into close races. In the aftermath, pollsters and journalists floated several hypotheses to explain why the polls overstated Democratic vote shares by an average of 5 percentage points. We will use Pollster's extensive database of 2014 polls, as well as other data sources, to empirically test a few of these hypotheses to determine if they hold true for all polls, hold true for only some or certain types of polls, or if the conjecture is unsupported. Specifically, we will address the following claims: (1) Polls captured the general trend toward Republicans between Labor Day and Election Day, but missed the magnitude of the shift that occurred in the final days of the campaigns. (2) Likely voter models and screens overestimated turnout, which possibly meant the nonvoters included in likely voter samples were more likely to say they supported Democratic candidates than Republicans. (3) Undecided voters split disproportionately in favor of the Republicans on Election Day. (4) In-state pollsters picked up the trends better than national organizations. We suspect that each of these contributed to the systematic error in the 2014 midterm polls, but this project will provide evidence regarding how much the polls were affected by each source.

Judging The Accuracy of Public Opinion Polls in Referendums

Thomas R. Marshall, *University of Texas at Arlington*

Most research on the accuracy of public opinion polls in elections deals with elections involving candidates. However, about half the American states and most cities also allow issue elections, authorized either by citizen petitions or by legislatures. Very little research is available on the accuracy of pre-election polls in issue elections, largely because interest groups involved in these elections typically keep their pre-election poll results private and because few other polls are conducted. This is unfortunate because many issue elections are on important topics of great public interest. Because of an unusual federal court order, many pre- and post-election polls conducted by the tobacco industry are now available on state and local elections involving either higher taxes on cigarettes or public restrictions on smoking, and dating from the mid-1970s until the present. Comparing these poll results to the actual election returns allows an examination of the accuracy of baseline, benchmark, tracking, exit, and post-election polling in these issue elections.

Measurements and Determinants of Polling Accuracy: Comparing Measures of Accuracy and Assessing Effects of Polling Practices

Jacob Sohlberg, *University of Gothenburg, Department of Political Science*

Mikael Gilljam, *University of Gothenburg, Department of Political Science*

This paper examines the performance of multiparty election polls, i.e. differences between polls and election results. We compare several measures of polling accuracy, ranging from the simple average absolute deviation in percentage points (Mosteller 1949) to the more recent multinomial accuracy measure (Arzheimer & Evans 2014). Extending a two-party accuracy measure (Martin, Traugott & Kennedy 2005) to multiparty systems, the multinomial measure has a number of superior features over the traditional measures. For example, it penalizes errors on smaller parties more harshly than measures of absolute deviations. In addition, it can readily be transformed into a dependent variable, enabling multivariate analysis when explaining polling accuracy. We study effects on polling performance of factors such as question wording, sample size, and sample type (probability vs. non-probability). The analysis is based on election polls from the four latest Swedish national elections (2002, 2006, 2010 and 2014).

Evaluation of Methods for Polling Third Party Candidates

Joseph W. Lenski, *Edison Research*

Randy Brown, *Edison Research*

The 2014 elections included many third party candidates who had direct or indirect influences on the outcomes of many races. Polling organizations were confronted with many choices in how to measure the strength of third party candidates in their pre-election polling. Some organizations chose to list all of the third party candidates, some organizations chose to list some of the third party candidates and some organization chose to list none of the third party candidates in their "horse race" questions. While many individuals and groups have evaluated pre-election polling for the accuracy of their estimates for the margins between the top two candidates, there has been very little evaluation of the accuracy of polls in measuring third party candidates. In this paper we evaluate the accuracy of polling estimates for over 100 third party candidates in over 1000 polls conducted in 2014 for statewide and U.S. House races. We evaluate the accuracy by polling organization, by survey method used, and by the method of handling third party candidates on the questionnaire. We also interview many survey practitioners about the challenges they face in making decisions about how to handle third party candidates and what information they are looking for in order to make better information methodological decisions. The paper will take the findings and make recommendations about how to handle third party candidates based upon the specific characteristics of each election.

What Happened in North Carolina? The 2014 Elections Through the Lens of the High Point University Poll

Martin J. Kifer, *High Point University*

Brian McDonald, *High Point University*

The 2014 U.S. Senate and House elections were in some ways a predictable midterm loss for a relatively unpopular president. By Election Day, most models showed the likelihood of a Republican takeover of the U.S. Senate to be relatively high. To a large extent, the predictive models were accurate. But much commentary after the votes were tallied focused on the manner in which public opinion polls reflected the final results. In North Carolina, in particular, there was discussion about why some polls just before the election diverged from the final

AAPOR 70th Annual Conference

result. The High Point University Survey Research Center fielded a total of five polls in three states (NC, CO, and NH) prior to Election Day. These live interviewer and automated surveys lend insight into the motivations of the voters in each of these states. Comparisons of the attitudes of voters seen through the surveys on a number of key metrics show high levels of agreement on some crucial issues, but also divergence on some as well. This project examines the topline and crosstab findings of each of these studies, compares their findings to other studies inside NC and in other battleground states, and applies multivariate analyses to the data to tease out additional findings. The project outlines a series of lessons learned from the 2014 elections that may be applied to future presidential and midterm elections.

The Most Important Problem: Immigration, Health or Energy?

Problems with the “Most Important Problem Question:” How Question Wording and Format Impact the American Public’s Priorities

Trevor Tompson, *The Associated Press-NORC Center for Public Affairs Research*

David Sterrett, *The Associated Press-NORC Center for Public Affairs Research*

Jennifer Benz, *The Associated Press-NORC Center for Public Affairs Research*

Dan Malato, *The Associated Press-NORC Center for Public Affairs Research*

Becky Reimer, *The Associated Press-NORC Center for Public Affairs Research*

Emily Alvarez, *The Associated Press-NORC Center for Public Affairs Research*

Nicole Willcoxon, *The Associated Press-NORC Center for Public Affairs Research*

One of the most commonly asked survey questions of the American public is: “What do you think is the most important problem facing this country today?” Many research and news organizations frequently ask this question, and they usually allow respondents to offer one or two answers. The public’s response to this question is highly publicized, and the media and policymakers use it to frame the political agenda. However, asking people to choose only one problem presents people with a false choice that oversimplifies the complexity of the modern political environment. An AP-NORC Center survey shows that the wording and answer format of such a question can have a significant effect on the results. On a nationally representative web survey of 1,141 Americans in December 2013, respondents were asked the following variation of the most important problem question: “Thinking about the problems facing the United States and the world today, which problems would you like the government to be working on in the year 2014?” Respondents were then given space to list up to ten problems and were asked how much effort the federal government should devote to solving each problem they identified. The results from this survey illustrate that most people identify more than one important problem and want policymakers to focus on solving multiple problems. Comparing the responses to this question with the responses to the standard most important problem question on surveys conducted during the same month shows that allowing respondents to give more than one answer can significantly change the outlook of public opinion and tells a different story about American priorities. The results confirm the common finding that the public is concerned about economic issues, but also illustrate that many people want the government to solve problems related to issues such as education or immigration.

AAPOR 70th Annual Conference

Contact, Contexts, and Public Attitudes Toward Illegal Immigration

Timothy B. Gravelle, *University of Essex*

Illegal immigration has long been a contentious issue on the American policy agenda. In endeavoring to understand the sources of mass public attitudes toward immigration, social scientists have drawn on theories of intergroup contact to argue that contact with immigrants shapes immigration attitudes. Absent direct measures of interpersonal contact, contextual measures such as respondents' racial/ethnic milieus or proximity to salient geographic features (such as borders) have been used as proxies of contact. Such a research strategy still leaves the question unanswered: is it contact or context that really matters? Drawing on nationally representative survey data from the Pew Research Center, this paper evaluates the predictive efficacy of both personal contact with illegal immigrants and contextual measures (county-level Hispanic population and proximity to the US–Mexico border) on white non-Hispanic Americans' attitudes toward illegal immigration. It finds that contextual factors are predictive of public attitudes; personal contact is not. These findings call into question the assumption that contextual measures serve as a straightforward proxy for interpersonal contact. The paper discusses some processes (other than interpersonal contact) through which contextual factors may influence public attitudes.

Beyond Innumeracy: Examining Qualitative Misperceptions About Immigrants in Finland

Daniel E. Herda, *Merrimack College*

Population innumeracy (the tendency for native born respondents to over-estimate the size of immigrant or minority populations) has sparked growing scholarly interest. Many fear that such widespread misperceptions will generate inter-group hostility. However, erroneous size estimates may not be the only consequential misperception. There are also qualitative questions that are prone to error, such as: What is the most common origin of immigrants? Using a representative sample from the Finnish National Election Survey, the current study provides the first examination of misperceptions about the primary source of immigration. I consider both their extent and correlates among a sample of Finnish citizens. Results indicate that about one-fifth are incorrect, with most faultily identifying Somalia as the most common source. These misperceptions are related significantly to age, gender, socio-economic status and media exposure, suggesting that they are not a result of random ignorance. Further, source misperceptions are strongly associated with increased perceptions of cultural threat from immigrants, suggesting potential consequences for such faulty views. Overall, the findings indicate that research on population innumeracy should expand to include misperceptions beyond size.

Visual Framing of Unconventional Energy: How Support Shapes Interpretations of Environmental Risk and Economic Opportunity

Erik P. Bucy, *Texas Tech University*

Amber Krause, *Texas Tech University*

Matthew Van Dyke, *Texas Tech University*

Andy King, *Texas Tech University*

Melanie Sarge, *Texas Tech University*

An important driver of the policy discussions and public opinion debates that surround controversial energy production technologies, especially hydraulic fracturing (or fracking), is the

AAPOR 70th Annual Conference

media framing used to represent the technology in the popular press. News framing consists of both verbal and visual elements within stories that emphasize particular narrative understandings and, in Entman's formulation, influence problem definitions, causal interpretations, moral evaluations, and treatment recommendations. This paper reports on an original online experiment designed to investigate how support for unconventional energy production technologies shapes interpretations of news images used to represent the fracking issue. For the study, 40 images (20 representing economic benefits and 20 representing environmental risks) drawn from news stories and information sites about fracking were shown to 250 participants (44% female; 48% aged 31-50; 70% white; 44% liberal, 31% moderate, 25% conservative) and were evaluated for their content focus and emotional intensity. Each participant was assigned to evaluate seven randomly selected images from both groupings and in an open-ended question were asked to describe their thoughts and feelings about each evaluated image. From an attitude questionnaire in the pre-experiment portion of the study, participants are grouped into supporters, opponents, and undecideds. The analysis focuses on interpretations of perceived economic benefits and environmental risks among these three groups, with particular attention paid to what moves undecideds (roughly 60% of the sample). Findings indicate that undecideds are more nuanced and ambivalent in their interpretation of the images overall, and see more positive and more negative elements in images that are viewed by supporters and opponents as either positive or negative. Implications for professional practice and subsequent research are discussed and follow-up studies suggested.

Surveying Children and Teens

A New Tool to Collect Ego-Centered Network Data in Online Surveys

Tobias H. Stark, *Utrecht University*

Jon Krosnick, *Stanford University*

Questions about survey respondents' social contacts (ego-centered networks) have become increasingly popular in surveys because research has shown that social contacts influence people's behavior and attitudes. How to collect information on a respondents' social network in online surveys is, however, not well understood. Questions about ego-centered networks impose a high cognitive burden on respondents because the same questions must be answered for all network contacts. In online surveys, where no interviewer can motivate the respondents, ego-centered network questions lead to high break-off rates and more item-nonresponse, and respondents tend to nominate fewer network contacts. This study presents a new data collection tool for ego-centered networks in online surveys. We make use of Web 2.0 graphical features to generate a more engaging experience for survey respondents when answering the same question for all of their network contacts. We will present the finished tool and report results from an experiment with an U.S. online convenience sample (N = 430), comparing the data quality of the new tool with a classical ego-centered online survey. We also report results from an experiment with Dutch school children (about 11 years old) in which the tool was used to collect information of the children's out-of-school friends. We hope that the interactive features of the survey tool allow collecting high-quality data also among this age group.

Luck of the Draw: A Comparison of Probability-Based Samples for Youth

Jennifer Cantrell Cantrell, *American Legacy Foundation*

Elizabeth C. Hair, *American Legacy Foundation*

Valerie Williams, *American Legacy Foundation*

Alexandria A. Smith, *American Legacy Foundation*

Randall K. Thomas, *GfK*

Donna M. Vallone, *American Legacy Foundation*

Probability-based samples are essential for making inferences about population parameters, including those related to health behaviors for surveillance and evaluation. There are different approaches to selecting probability-based samples, each with varying costs and efficiencies. This study compared tobacco use estimates across three probability-based samples recruited via varying methods. From May to August 2014, we used 3 probability-based samples to recruit 15 to 21 year old individuals for a 3-year longitudinal study of tobacco use and tobacco-reduction advertising. The first sample used an existing online sampling frame (GfK's KnowledgePanel®) to recruit a probability-based sample of 1,835 individuals. The second sample relied on address-based sampling (ABS) methodology to recruit a random sample of 10,257 individuals. The third sample employed a variant of dual-frame RDD which included a targeted cell phone sample to recruit 1,966 individuals. Each sample used the same web-based survey instrument for data collection and the resulting data were weighted to the same set of target benchmarks, after adjusting for selection probabilities. For key surveillance measures, the weighted estimates were comparable when obtained from the KnowledgePanel and ABS samples. However, the corresponding estimates from the telephone recruitment methodology were significantly different from the other two. In addition, the unequal weighting effect was much higher for the phone methodology. While there have been differences found for population estimates in different national surveys, very few have employed two or more alternative probability sampling methods in the same study to compare their estimates and potential biases. In this study, we found both similarities and differences in a number of the behavioral measures and the associated attitudes. In addition to discussing these findings, we also will explore the impact of weighting in the method differences we observed.

Does Modality Matter? Comparing Smoking Behavior Among 18-21 Year Olds Across Online, In-Person and Telephone Surveys

Peter Messeri, *Mailman School of Public Health, Columbia University*

Jennifer Cantrell, *Legacy Foundation*

Valerie Williams, *Legacy Foundation*

Haijun Xiao, *Legacy Foundation*

Alexandria Smith, *Legacy Foundation*

Paul Mowery, *Biostatistics, Inc*

Donna Vallone, *Legacy Foundation*

The online survey, as the most recent phase in survey research technology, raises perennial concerns regarding how changes in technology influence population-based sample estimates of health behaviors. We compared sample estimates of smoking behavior among 18-21 year olds from a 2014 national online survey with two other national online surveys, two national surveys using computer-assisted in-person interviews and two that were a mix of land line and cell phone surveys. Point estimates of lifetime smoking across all surveys ranged from 20% to 55%; with the online survey estimates forming a relatively tight cluster in the middle of this range from 29% to 37%. Current smoking (within 30 days of interview) ranged across all surveys from 13%

AAPOR 70th Annual Conference

to 31%. The online surveys formed a tight cluster of estimates at the lower end of this range between 13% and 19%. Mode effects were not statistically significant for either the lifetime or current smoking estimates. We next compared the range of sample coefficients from regression equations that included as independent variables gender, race/ethnicity (Non-Hispanic White, Non-Hispanic Black, Hispanic, other) and education (in college, H.S. graduate not in college, less than H.S.). The gender and race/ethnicity sample coefficients for online surveys were not consistent with those for the in-person and telephone surveys. The sample regression coefficients for education were consistent across modes with the highest rates of lifetime and current smoking among non-H.S. graduates. Modality did matter when estimating smoking behavior, but not in a consistent way. Closer inspection of these surveys indicated that along with changing technology there were differences in question wording across surveys that may have contributed to the modality effects. These findings highlight complex changes that occur with shifting technology that merit continued research into modality differences as we move deeper into the online era of survey research.

Implementing a Large Scale Population-Based Youth Online Health Survey in a Non-School Setting

Amy Kristen Leite Bennett, *Hennepin County Public Health Department*

Mei Ding, *Hennepin County Public Health Department*

David C. Johnson, *Hennepin County Public Health Department*

Urban Landreman, *Hennepin County Public Health Department*

Komal Mehrotra, *Hennepin County Public Health Department*

Emily Thompson, *Hennepin County Public Health Department*

Data collection directly from adolescents is optimal for estimating prevalence of health behaviors. Few studies were found using population-based data collection methods among adolescents in non-school settings. This project examines the implementation of an online population-based survey of youth age 13 to 17 in a non-school setting in Hennepin County, Minnesota. A major feature was obtaining Institutional Review Board approval and identifying strategies to obtain parental consent while ensuring privacy for adolescent respondents. In January 2015, adolescents (aged 13-17) living in households selected to participate in another survey (Child SHAPE) will be invited to participate in the password-protected online Youth SHAPE survey. The methodology uses four mailings to encourage online response, including: a pre-notification postcard, a letter which includes unique username and password for each respondent, a reminder post card, and a second and final letter. A stratified sample was drawn using a U.S. Postal Service Delivery Sequence File. Twenty-six thousand households were selected using three sets of stratification: geographic strata, oversampled areas, and "child flag." The DSF was stratified into 2 geographic areas, and further divided into oversampled areas, identified as census tracts where hard-to-reach groups were likely to be clustered based on race/ethnicity and poverty. In addition, a child presence flag was developed using state and county program records (e.g. birth records), indicating the likely presence of a child in the household. Oversampled and "child flag" tracts were sampled at higher rates compared to other tracts. The presentation will discuss: 1) response rate for the Youth SHAPE 2015 survey relative to parental participation to the Child SHAPE survey; 2) implications of IRB approval in obtaining parental consent; 3) effectiveness of "child flag" in identifying households with children compared to non-flag households; and 4) lessons learned in administering online surveys to minor youth in a non-school setting.

Discrimination and Digital Abuse Among Socially Networked Teenagers and Young Adults in the United States

Emily R. Alvarez, *The Associated Press-NORC Center for Public Affairs Research*
Jennifer Benz, *The Associated Press-NORC Center for Public Affairs Research*
Trevor Tompson, *The Associated Press-NORC Center for Public Affairs Research*
Jennifer Agiesta, *The Associated Press*

The prevalence of digital communications and networking through social media among teenagers and young adults has reshaped young people's social behavior in recent years. A growing body of research suggests that young people frequently encounter discriminatory language online and frequently experience digital abuse. A new study by MTV and The Associated Press-NORC Center for Public Affairs Research explored the pervasiveness of these experiences among teens and young adults, how they are affecting America's youth, and how young people are responding to these experiences. We conducted national surveys of teenagers and young adults between the ages of 14 and 24 in 2009, 2011, and 2013 to measure the exposure of young people in the United States to online discriminatory language and to better understand where on the Internet young people encounter these messages, as well as to assess young people's experiences with online digital abuse. The latest findings suggest that while the use of discriminatory language online remains commonplace, a clear majority now say it is never okay to use this type of language, even when joking among friends. Exposure to discriminatory language and images varies significantly by social networking site, with YouTube, gaming communities, and Facebook among the most commonly cited platforms in which discriminatory language and images are encountered. Further, a significant number of teens and young adults are involved in or are affected by digital abuse, and nearly three-quarters of young people say digital abuse is a big problem for society that needs to be addressed. This study explores the challenges of asking sensitive questions to a young population and tracking trends in an area where the technology and language changes regularly.

Testing Alternative Methods to Enhance the Validity and Feasibility of Internet Based Research Among Children and Their Parents

Keisha Miles, *Institute for Survey Research - Temple University*
Rachel Teneralli, *The Children's Hospital of Philadelphia*
Heidi Gunwald, *Institute for Survey Research - Temple University*
Nina Hoe, *Institute for Survey Research - Temple University*

BACKGROUND: The rapid advance in technology within the last decade has fostered the rise in accessibility to and use of the internet among U.S. households. As a result, the internet has become a viable and valuable medium among health researchers today. Despite the many associated benefits with online data collection methods, a significant concern remains over the validity of such methods for child research. Researchers are unable to guarantee children are completing the materials on their own behalf therefore uncovering novel techniques that help validate the data being collected are needed. This study sought to test an interactive voice recording (IVR) survey as novel technique to assess help test the reliability and validity of online survey data for young children. **METHODS:** An IVR questionnaire was added to a large cross-sectional study collecting patient reported outcome (PRO) questionnaire amongst children aged 8-17 years and their parents (parent child-dyads) through an online research company. Upon completion of the online study questionnaire, parent participants were instructed to call to complete a short IVR phone questionnaire consisting of 10-questions; 5 questions answered by

AAPOR 70th Annual Conference

the parent and 5 questions by the child. Each IVR questionnaire was recorded and transcribed. Verification both parent and child answered was conducted by IVR system and through manual coding by trained research staff (2 per recording). Any discrepancies were adjudicated.

RESULTS: 4,005 dyads completed the online PRO questionnaire and 1,623 (40.5%) called to complete the IVR questionnaire. 76% of dyads (n=1,237) answering the questionnaire had both child and adult voices verified. CONCLUSIONS: Although the majority of cases had both child and parent voices verified completion rates were lower than expected. Increase compliance and using mixed methods analysis may further support the use of IVR as a valid method to validate online data in for young children.

Field Effects, Survey Participation and Data Quality

Evaluating Two Differential Survey Treatment Strategies

Gerry Dirks, *Experian Marketing Services*

Karen Swift, *Experian Marketing Services*

Pat Pellegrini, *Experian Marketing Services*

Charlie Palit, *University of Wisconsin*

Christine Kudisch, *Experian Marketing Services*

David Lustig, *Experian Marketing Services*

Researchers have used differential survey treatments (DST) for many years. This paper describes two different DST strategies to improve response rates. The first strategy involves appending auxiliary data to sample in order to attract hard-to-reach populations in the first phase of a two phase recruitment process. Hard-to-reach populations here are defined as households with a Hispanic or African American presence, households that contain adults 18-34 and households with no land-line telephone. We evaluate the effectiveness of auxiliary information as well as differential incentives within a standard experimental design. We describe the effectiveness of this strategy across a number of hard-to-reach populations as well as speculate on its potential effects on non-response bias. The second DST strategy outlines a newer, more complex approach to DST that involves constructing a multivariate model that takes into account not only historical response rates by demographic and geographic variables but also the timing of their response to a phase 1 consent phase as well sample type. Late responders to phase 1 are primed with higher incentives. Sample type also plays a part – respondents from fresh sample will on average receive less than respondents from previously worked sample. The multivariate model predicts difficulty scores for each respondent in the sample based upon the criteria outlined above. Difficulty scores are then aggregated into difficulty classes and an incentive is assigned to that class. The objective is to develop a dynamic DST that simultaneously takes into account response behavior from a number of demographic and geographic groups as well as other variables such as phase 1 behavior and develop incentives classes that reflect this analysis. This approach can be utilized as a static model applied once to a sample or as a dynamic model applied with real-time to a constantly changing sample pool.

Fieldwork Effort, Response Rate and the Distribution of Survey Outcomes: a Multi-Level Meta-Analysis

Patrick Sturgis, *University of Southampton*

Ian Brunton-Smith, *University of Surrey*

Joel Williams, *TNS-BMRB*

As fieldwork agencies devote ever greater resources to mitigate falling response rates in face-to-face interview surveys, the need to better understand the relationship between level of effort, response rate, and nonresponse bias grows ever more pressing. In this study we assess how response rates and outcome distributions change over the number of calls made to a household. Our approach is comprehensive rather than selective: we analyse change in the response distribution over repeated calls for over 700 survey variables, across seven different major surveys in the UK. The four surveys cover different topic areas and have response rates which vary between 54% and 76%. Comparisons are made for both unweighted and post-stratified estimates. We code each question on a number of different attribute dimensions to produce a broad typology of question types and then analyse nonresponse bias (defined as the difference between the point estimate at call *n* and the final response distribution for the full sample) within a multi-level meta-analytic framework, where estimates of bias are nested within calls and within questions, and questions are nested within surveys. This approach enables us to model how estimated bias varies systematically as a function of call number (fieldwork effort), question type, and survey topic as well as interactions between these characteristics. In addition to contributing to our understanding of how effective fieldwork effort is in reducing nonresponse bias, our study also includes an assessment of the cost-effectiveness of additional fieldwork effort at different points in the fieldwork cycle.

Is it Worth the Effort? Contact Attempts and Nonresponse Bias Reduction in a Large-Scale Cross-National Survey

Tanja Kunz, *Darmstadt University of Technology*

Marek Fuchs, *Darmstadt University of Technology*

Response rates have been declining over the years. In order to compensate for potential nonresponse bias, high quality academic and government surveys have adapted and further developed their field work procedures. In recent years, it has been suggested to use adaptive designs (Wagner, 2008) or responsive designs (Peytchev et al., 2010) in order to target sample units belonging to underrepresented groups. However, many large-scale surveys still adopt a more general field work strategy. Often, the overall number of contact attempts has been increased in order to reduce nonresponse and nonresponse bias. Several studies have demonstrated (Heerwegh et al., 2007; Kreuter et al., 2014) that an increasing number of contact attempts helps boost response rates. Nevertheless, even after multiple contact attempts post-stratification and raking procedures are deemed necessary to compensate for nonresponse bias because increasing response rates do not guarantee a linear decline in nonresponse bias. However, multiple contacts may attract already overrepresented groups and thus nonresponse bias may remain on a stable level or even intensify. Consequently, researchers have to decide whether additional contact attempts actually pay off in terms of nonresponse bias reduction. This paper is based on data from the European Social Survey (ESS), a biennial face-to-face survey in the general population of more than 30 participating countries. Previous analyses using socio-demographic variables (Fuchs et al., 2013) provided preliminary evidence that additional contact attempts generally increase response rates, but at the same time, also have the potential to increase nonresponse bias. In this paper, we assess the effects of multiple contact attempts on nonresponse bias for a set of substantive variables. We will test whether

AAPOR 70th Annual Conference

and to what extent elevated response rates actually contribute to a reduction of nonresponse bias in attitudinal variables.

The Workload, Effort, and Quality Associated with Collecting Data on Vacant American Community Survey Addresses

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

The American Community Survey (ACS) requires data be collected on vacant sample addresses. Information such as why the unit is vacant (e.g. for rent, sale, or recreational use) may be collected from a respondent who indicates such on the internet or a mail questionnaire, over the phone, or during an interview with a computer-assisted personal interviewing (CAPI) interviewer. Because the ACS first invites sample cases to self-respond by an invitation mailed to their address, most vacant units are not classified until the end of the field period by face-to-face interviewers, the mostly costly ACS operation. Currently, over 25 percent of the ACS CAPI workload is determined to be vacant, 12% of which take more than three contact attempts to obtain a final outcome (Griffin & Nelson, 2014). Using data from nine months of 2013 ACS data collection, I discuss the effort applied to interviewing approximately 100,000 sample addresses classified as vacant. I describe the unweighted workload associated with ACS vacant and temporarily occupied interviews across each ACS data collection operation and present the number of mailings, CATI calls, and CAPI contact attempts made to such addresses. I also estimate the cost of CAPI vacant interviews and contact attempts. I then evaluate the quality of responses obtained for vacant housing units during nine months in 2013. I discuss the reliability of vacant address classification, document item nonresponse rates for vacant sample addresses, and report on the degree to which ACS relies on CAPI interviewers to classify sample addresses as vacant units by “observation only.” I conclude by discussing future research designed to (1) improve the quality of classifying vacant housing units, (2) increase the quality of data obtained on such addresses, and (3) decrease the cost of efforts related to interviewing “no one.”

Towards Determining an Optimal Contact Attempt Threshold for a Large-Scale Personal Visit Survey

Adam Safir, *U.S. Bureau of Labor Statistics*

Lucilla Tan, *U.S. Bureau of Labor Statistics*

As with other federal statistical agencies, the Bureau of Labor Statistics continues to operate under budget constraints. While improving the design of the survey instrument and increasing the options for more convenient and/or efficient data capture are two important ways for maximizing returns on the survey operations budget, there is an additional dimension in personal-visit surveys for discretionary action that can make a difference – setting a maximum threshold for the number of contact attempts, or the level of effort, expended to resolve a case. In particular, the cost of contact attempts can be significant for personal-visit surveys with eligible sample units spread over an expansive geographic area. Before fielding a large-sample test of a proposed “optimal” contact attempt threshold, the authors initiated this study as a verification of findings from an earlier study that recommended seven attempts as the threshold for resolving a sample unit’s final disposition. Using more recent data and additional metrics, the authors perform a retrospective analysis of first wave panel data from the Consumer Expenditure Interview Survey (CEQ) collected between April 2012 and March 2014. The results to be reported on include impact on pre-survey reluctance, response rate, reporting quality, and potential cost savings; and plans to employ R-indicators to examine sample representativeness and to estimate the error bounds of the key evaluation metrics via bootstrap resampling. The

findings will be of interest to survey methodologists and practitioners working in large scale survey operations with constrained budgets and specific data quality goals.

Multi-Mode Surveys Minimizing Cost While Maintaining Quality

Telephone Prompting to Obtain Survey Participation via Less Expensive Modes: Results from an Experiment in a Nationally Representative Mixed-Mode Establishment Survey

Lauren D. Harris-Kojetin, *National Center for Health Statistics*

Manisha Sengupta, *National Center for Health Statistics*

Melissa Hobbs, *RTI International*

Angela Greene, *RTI International*

Vincent Rome, *National Center for Health Statistics*

Decreasing response rates and increasing costs are enduring survey challenges. We report results of a randomized experiment embedded in the National Study of Long-Term Care Providers (NSLTCP) sponsored by the National Center for Health Statistics. The biennial NSLTCP uses a multi-mode design, with mail and web questionnaire options offered first (via three mailings) followed by telephone interview follow-up attempts to non-respondents. During the inaugural 2012 wave, some of the sample respondents did not respond to the questionnaire mailings, thus needing telephone interview follow-up; others called the assistance line for help answering survey questions or accessing the Web questionnaire, or to report being out-of-business. The 2012 response via mail and Web prior to telephone interview follow-up was lower than expected; more cases than anticipated required telephone interview follow-up, thus increasing costs. For the 2014 wave, we tested the effectiveness of prompting calls in driving a greater number of mail and Web completes earlier, thus decreasing the number of cases needing telephone interview follow-up. The purpose of the prompting calls was to: 1) confirm the mailing was received; 2) encourage participation by mail or Web; 3) address concerns about participating; 4) answer questions about questionnaire items; 5) obtain new contact information if the mailing was not received; 6) provide technical assistance to access the Web questionnaire; and 7) identify out-of-business cases before telephone interview follow-up. Among cases not responding within three weeks of the first mailing, 2,000 were randomly assigned either to be telephone prompted or not. Preliminary results suggest that before telephone interview follow-up started, a higher percentage of prompted cases than non-prompted cases had eligibility status determined (35% versus 27%) and a lower percentage were still pending (64% versus 73%). Final outcomes (whether eligibility status determined, return rate, item missing or data retrieval needed for returned questionnaires) will be presented.

Knowing When to Stop: Evaluating First 5 LA Family Survey Data Based on Data Collection Mode and Difficulty to Complete an Interview

Andrea Mayfield, *NORC at the University of Chicago*

Ying Li, *NORC at the University of Chicago*

Alicia Frasier, *NORC at the University of Chicago*

Ned English, *NORC at the University of Chicago*

Jeni Vanicek, *NORC at the University of Chicago*

Joelle Greene, *Harder+Company Community Research*

Melinda Leidy, *First 5 LA*

Multi-mode surveys targeting hard-to-reach populations are challenged with efficiently identifying and interviewing eligible respondents. Key decisions include how extensively to pursue less-cooperative respondents and through which data collection modes. The First 5 LA Family Survey was developed to provide representative data about key indicators of well-being for children ages zero through five across 14 communities in LA County, CA. The study employed a multi-mode address-based sample (ABS) design with the aim of efficiently reaching households with young children and achieving ambitious data collection targets. The design for the survey began by sending all households a self-administered questionnaire (SAQ) screener which requested information about the number of age-eligible children living in the household, household telephone number, and familiarity with First 5 LA. Eligible households that returned an SAQ screener with a phone number were contacted by telephone. Eligible households that did not provide a phone number were contacted in-person. We attempted to match telephone numbers to addresses for households that did not return an SAQ screener for follow-up via telephone interviewing; in-person follow-up was attempted with any non-responding households not matching to a phone number. In-person interviewing proved to be the most effective mechanism to yield completed interviews, but also incurred the highest cost. This paper seeks to answer the question: how, if at all, did key survey estimates change by including in-person interviews? Could similar results have been achieved with less extensive follow-up and less expensive data collection modes? This paper compares easy-to-complete cases (e.g., those who returned a screener with a telephone number) to less easy-to-complete cases (e.g., those who returned a screener without a telephone number) and difficult-to-complete cases (e.g., non-respondents who required follow-up) using bivariate analysis and regression. Combinations of these categories will also be compared to determine how adding follow-up modes may have changed key estimates.

Benefits and Challenges of Web Surveys in Mix-Mode Designs: Demographic and Data Quality Differences Across Modes in Survey of Households Recovering From Superstorm Sandy

David Sterrett, *Associated Press-NORC Center for Public Affairs Research*

Dan Malato, *Associated Press-NORC Center for Public Affairs Research*

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

Trevor Thompson, *Associated Press-NORC Center for Public Affairs Research*

Jennifer Benz, *Associated Press-NORC Center for Public Affairs Research*

Becky Reimer, *Associated Press-NORC Center for Public Affairs Research*

Due to cost and coverage challenges, many research organizations are now conducting surveys via the web as part of mixed-mode designs. The increasing use of the web in combination with modes such as telephone and face-to-face raises several important research questions. Are there differences in respondents who opt to complete the survey in each mode? Are there

AAPOR 70th Annual Conference

differences in response quality across modes? A recent AP-NORC Center survey of how 12 neighborhoods in New York and New Jersey are recovering from Superstorm Sandy provides an opportunity to explore such questions about mixed-mode surveys. All households in the address-based sample were first sent an advance letter directing them to a URL to complete the survey. After a two-week delay, cases with matched landline telephone numbers were dialed, and a sample of cases without matched telephone numbers were contacted by field interviewers. Reminder mailings were sent to a randomly selected set of non-respondents with \$2 or \$5 pre-paid incentives. Of the 1,009 completed interviews, 316 were via phone interviews, 393 were via face-to-face interviews, and 300 were via web self-administered questionnaires. Although respondents were not randomly assigned to a mode, the surveys were identical across mode and all of the responses were collected between June 28 and September 9, 2014. The analysis shows that there are clear demographic differences between respondents who opted to complete the survey via the web than those who completed the survey with phone or field interviewers. Measures of data quality such as breakoffs, item nonresponse and straight-lining are relatively similar across the three modes. However, the survey results show that the web can reduce social desirability bias among respondents. The findings from this study provide researchers with insights into the potential benefits and challenges of including web surveys in mixed-mode designs.

Mixed-Mode Experiment - Evaluation of Effects on Data Quality, Response Rates and Cost Reduction

Mikaela Johanna Jarnbert, *Statistics Sweden*

Johan Eklund, *Statistics Sweden*

Over recent decades, response rates in general surveys have decreased rapidly in Sweden as in many other countries. This has led to higher uncertainty in estimates as well as higher costs of data collection. One way to deal with this problem is to allow other response modes. In September 2014, Statistics Sweden carried out an experiment with mixed-mode (telephone interviews and web questionnaires) within the Swedish Party Preference Survey. The survey sample is a probability sample which is drawn from the Swedish population register and the sample was randomly assigned to web questionnaire (experiment group) or telephone interview (control group). The experiment group received the opportunity to respond on the web at first but after one week of field work this group was processed simultaneous web/interview. In order to acquire control over the data collection a specific tool was developed in order to combine the telephone and web instruments enabling smooth changes of type of mode and deregistration of sample units. In this paper we set out to study whether or not the cost of data collection decreased, whether or not the response levels increased and if the mixing of modes lead decreased data quality or mode effect. The evaluation of the experiment is done using data set of Party Preference Survey together with information from national population registers and para data from the data collection (from telephone system as well as web system). The evaluation of the experiment is facilitated by both the panel design of the Party Preference survey, which allows us to study the same individuals over four waves of the survey, and information from register for both respondents and non-respondents. Taking all this together this experiment is unique for studies of mode effects, effects on data quality, response rates, nonresponse errors and possible cost reduction.

Panel: Twitter Data for Social Research: Empirical Evidence

Panel: Twitter Data for Social Research: Empirical Evidence

Yuli Patrick Hsieh, *RTI International*

Joe Murphy, *RTI International*

Data from social media sites and particularly Twitter are increasingly being leveraged to answer social research questions that were once the domain of surveys. As we enter into the new world of big social data, it is imperative that we critically examine what data from Twitter can and cannot do for us. This panel presents a collection of papers designed to empirically explore how Twitter data relate to phenomena of interest for social researchers. We hope that these examinations begin to reveal some of the purposes for which Twitter data might be used either in conjunction with or in place of traditional surveys. Papers in this panel simultaneously illustrate the breadth of potential applicability for big data methods as well as constraints on the ability for large-scale unrepresentative social data to inform our understandings. In many ways, the panel builds on the recent release of the AAPOR reports on Social Media in Public Opinion Research and Non-probability sampling. Each of the empirical examples presented integrates theoretical considerations for the use of big social data with a critical empirical assessment of what those data imply about the social processes of interest. In this manner, we hope to begin to demarcate when these sorts of methods will yield actionable understandings of society and when they are not “fit for purpose.”

Total Twitter Error? A Discussion of Surveys and Twitter for Examining Attitudes toward Marijuana Legalization

Yuli Patrick Hsieh, *RTI International*

Joe Murphy, *RTI International*

With the continued decline in survey response rates and rise of social media as a popular means of communication, researchers are investigating the latter as source of information on opinions and behaviors traditionally measured through the former. Twitter, in particular, has been used increasingly capture sentiment or discussion without any interaction with the “respondent.” In some ways, Twitter can be viewed as a large opt-in “survey” with no control (or influence) of the researcher regarding question wording or reporting format from a respondent. Several studies have claimed to replicate survey findings using Twitter data, suggesting a potential to supplant designed survey data with organic big data. As the central organizing structure of the field of survey methodology, the Total Survey Error (TSE) framework can provide a starting point for assessing Twitter as a research resource. For example, analysis of tweets is subject to both observation and non-observation errors. However, TSE will not sufficiently capture all areas of concern with Twitter, nor will all components of TSE be applicable to Twitter. In this presentation, we will dissect the TSE approach and its applicability to Twitter focusing on the issue of measuring attitudes toward marijuana legalization. We begin by discussing the trends of measurement of these attitudes from existing national surveys, presenting what is known about the real or potential error sources. We then examine attitudes towards legalization shared on Twitter using incrementally refined search query specifications. We relate the relative strengths and weaknesses of these queries to the TSE notions of coverage and measurement error. Our goal is to shed light on what can and cannot be compared between surveys and social media. We conclude with reflections on the viability of TSE framework for social media data and thoughts on the role of social media in survey measurement moving forward.

Using Twitter Data to Calibrate Retrospective Assessments in Surveys

Josh Pasek, *University of Michigan*

Elizabeth Hou, *University of Michigan*

Michael F. Schober, *New School for Social Research*

Frederick G. Conrad, *University of Michigan*

Cliff Lampe, *University of Michigan*

Lauren Guggenheim, *University of Michigan*

Retrospective survey questions are important for many social measurements, ranging from economic indicators to health tracking. Yet a long line of research in survey methodology has shown that respondents do not always answer these questions in the ways researchers intend. When responding to some questions, individuals tend to “telescope”—erroneously including events from before the time period referenced in the question. This study explores the use of Twitter data to determine how respondents appear to construct retrospective evaluations of the economy. Specifically, we use cross-correlations to link self-reported measures of economic performance in different time periods, as assessed by the Survey of Consumer Attitudes, with sentiment in Tweets containing the word “jobs” from 2008-2014. This measure of sentiment is used because it was found to be highly predictive in an earlier study (O’Connor et al., 2010). The use of cross-correlations allows us to identify the lags where correspondence between survey and Twitter measures is at its strongest. Instead of a momentary spike in Tweets’ predictive power at precisely the lag corresponding to the survey question’s reference period, we instead find that Twitter data from a large range of time points predict answers to retrospective survey questions about the economy in the past year. This pattern is not apparent for contemporary economic evaluations, where the corresponding time period is short. The results imply that individuals think about a range of time points when making retrospective economic judgments, what we term a temporal latitude of acceptance. Relevant time periods tend to be generally more recent than the 12-month window that respondents are requested to use in the survey, suggesting that respondents do not have a consistent temporal reference point and generally think of periods shorter than those queried for the questions examined. Tweets may help quantify telescoping error in surveys.

A "Collective-vs-Self" Hypothesis for When Twitter and Survey Data Tell the Same Story

Frederick G Conrad, *University of Michigan*

Michael F. Schober, *New School for Social Research*

Josh Pasek, *University of Michigan*

Lauren Guggenheim, *University of Michigan*

Cliff Lampe, *University of Michigan*

Elizabeth Hou, *University of Michigan*

Are there principled explanations for when a collection of tweets, which is inherently non-representative, will yield conclusions similar to those of survey data from a representative sample? We investigate a “collective-vs-self” hypothesis, predicting more likely alignment between estimates based on these two sources of information when the survey questions to which the tweets are compared concern groups larger than individuals or their households. We analyzed alignment between data from the Survey of Consumers (SCA) and sentiment analyses of tweets containing the word “jobs” from 2008-2014, and observed evidence in support of this hypothesis: Twitter sentiment frequently aligned better with a question about the

AAPOR 70th Annual Conference

collective ("Now turning to business conditions in the country as a whole--do you think that during the next twelve months we'll have good times financially, or bad times, or what?") than with the similar SCA question about "you and your family." We propose two complementary mechanisms through which tweets may reflect a broader trend than just the thinking of the individual user. First, tweets reflect posters' judgments of what their Twitter audience is interested in hearing about right now, which at some level reflects a summary judgment of what is on the minds of their followers and in the public discourse. Second, to the extent that ideas in tweets survive and flourish (i.e., are endorsed, retweeted, responded to, and lead to new followers), the poster's judgment that others are also thinking about, or likely to react to, this content is supported. We propose that although Twitter users do not represent the whole population, a corpus of tweets may cover the opinions and experiences of the broader population to the extent that it reflects a broader community and network-wide conversations.

Opportunities of Social Media in Personal and Societal Wellbeing

Munmun De Choudhury, *Georgia Tech*

Scott Counts, *Microsoft Research*

Eric Horvitz, *Microsoft Research*

Andrés Monroy-Hernández, *Microsoft Research*

Gloria Mark, *University of California, Irvine*

People are increasingly adopting social media platforms, such as Twitter and Facebook, to share their thoughts and opinions with their contacts. In a way, social media has transformed traditional methods of communication by allowing instantaneous and interactive sharing of information created and controlled by individuals, groups, and organizations: one in six people in the world today is a user of Facebook. An important attribute of social media is that postings on these sites are made in a naturalistic setting and in the course of daily activities and happenings. As such, social media provides a means for capturing behavioral attributes that are relevant to an individual's thinking, mood, communication, activities, and socialization. With the increasing uptake of social sites, there has been a corresponding surge of interest in utilizing continuing streams of evidence from social media on posting activity to measure and elucidate core aspects of human behavior that have traditionally been challenging. Particularly, through this panel, I would highlight our research on the harnessing of social media in reasoning about behavioral health concerns experienced by populations. Particularly, I will focus on behavioral measurement from two different scope and scale: microscopic or individual level and macroscopic or population level. For the first, I will cover findings on analyses and computational models that make automated inferences about the status and dynamics of postpartum depression in new mothers via postings made on Twitter and Facebook. For the second, I will focus on a study that examined the affective responses in Twitter experienced by communities in Mexico exposed to protracted armed conflict and how they might indicate desensitization to violence. Broadly, I will highlight the potential opportunities and challenges in use of social media as a novel stream of information for augmenting traditional approaches of health assessment.

Saturday, May 16
10:00 a.m. – 11:30 a.m.
Concurrent Session H

Mini-Conference: Matching and Weighting Opt-In Panels

Matching an Internet Panel Sample of Health Care Personnel to a Probability Sample

Charles DiSogra, *Abt SRBI*

Stacie Greby, *Centers for Disease Control and Prevention*

K.P. Srinath, *Abt SRBI*

Andrew Burkey, *Abt SRBI*

Carla Black, *Centers for Disease Control and Prevention*

John Sokolowski, *Abt SRBI*

Xin Yue, *Centers for Disease Control and Prevention*

Sarah Ball, *Abt Associates*

Sara Donahue, *Abt Associates*

Non-probability panels of volunteers can provide access to special or hard to reach populations that may not be easily reached using a probability sample; however, it is not generally appropriate to make statements regarding the precision of estimates based on a non-probability sample. Methods to calculate measures of precision for estimates measured by non-probability samples are necessary to improve the public health programs. The Centers for Disease Control and Prevention (CDC) currently conducts rapid surveys during and just after the influenza season each year to monitor vaccination coverage among health care personnel (HCP). These surveys use two non-probability internet panels for the sample, one opt-in panel of HCP professionals (e.g., physicians, nurses) and another general population opt-in panel. The purpose is to provide rapid information about influenza vaccination coverage plus related knowledge, attitudes, beliefs and behaviors among HCP. Using a non-probability sample is deemed fit for this purpose and application. The vaccination coverage estimates are published without measures of precision. We examined procedures to ex post facto match the 2013 non-probability panel samples of HCP with a probability-based sample derived from the 2013 National Health Interview Survey (NHIS) to improve the ability to assess precision and bias of the non-probability vaccination estimates. The steps used to align health care occupations between NHIS and panel HCP samples for information reported for a similar time window are presented. Vaccination coverage estimates were calculated using weights for both samples raked to identical population control totals. Variables are identified that contribute to observed differences between panel and NHIS estimates. We demonstrate matching a sample from the non-probability panel to the probability sample of HCP using a set of variables that will make the non-probability sample similar to the probability sample. A likely precision of vaccination coverage estimates using this matched sample is calculated.

Weighting and Sample Matching Effects for an Online Sample

J. Michael Brick, *Westat*

Jon Cohen, *SurveyMonkey*

Sarah Cho, *SurveyMonkey*

Scott Keeter, *Pew Research*

Kyley McGeeney, *Pew Research*

Nancy Mathiowetz, *University of Wisconsin, Milwaukee*

In September of 2014, we conducted coordinated data collections using an ABS probability mail survey (Westat), a probability web survey (Pew Research Center), and a non-probability web survey using SurveyMonkey's Audience panel (SurveyMonkey). A separate paper compares weighted estimates from these surveys, including comparisons to national benchmark estimates. This paper examines two statistical approaches that have been proposed to improve the quality of the estimates from non-probability samples. The first method involves using different weighting. Most weighting strategies have examined approaches intended to address coverage loss associated with lack of access or use of the internet. Propensity score weighting is one such approach. We explore alternative calibration weights, informed by data from the probability-based Westat and Pew surveys. The second approach is sample matching, where the respondents for the non-probability web survey are selected to match to targets from other samples or censuses. Since sample matching was not implemented in the SurveyMonkey web survey, we simulate this option by dropping observations from the data set so that the unweighted sample is more similar to observations from the probability mail survey. We explore the effects of both of these techniques by comparing the web survey estimates with and without the weighting and matching adjustments to the other probability sample estimates and benchmarks discussed above.

Can Surveys Posted on Government Websites Give Fair Representations of Public Opinion?

Michelle Kobayashi, *National Research Center, Inc.*

Non probability samples are in unrelenting ascendance among data collection designs no less so among local governments where questions are posted on websites with the intention of learning what the public thinks about important community issues. Socio-demographic weighting alone has failed to improve sufficiently the match of opt-in results with results from probability samples so the ability to generalize findings to the entire community is suspect. Using matched pairs of mail probability samples and opt-in surveys in eleven U.S. communities, National Research Center, Inc. examines a variety of weighting variables and techniques to compare results and determine if opt-in surveys can be substituted for probability samples when asking residents about the quality of community life, service delivery or public trust. Comparisons are made of unweighted data and post-stratification weighted data using resident socio-demographics augmented by then including behaviors related to civic engagement using voter turnout and area-specific responses to the Census's Current Population Survey.

Matching an Internet Panel Sample of Pregnant Women to a Probability Sample

Andrew M. Burkey, *Abt SRBI*

Charles DiSogra, *Abt SRBI*

Stacie Greby, *Centers for Disease Control and Prevention*

K.P. Srinath, *Abt SRBI*

Carla Black, *Centers for Disease Control and Prevention*

John Sokolowski, *Abt SRBI*

Helen Ding, *Centers for Disease Control and Prevention*

Sarah Ball, *Abt Associates*

Sara Donahue, *Abt Associates*

The Centers for Disease Control and Prevention (CDC) currently conducts surveys during and just after the influenza season each year to monitor influenza vaccination coverage among pregnant women (PW). These surveys use a non-probability Internet panel for the sample. The purpose and rationale is to provide affordable, timely information about vaccination coverage plus related knowledge, attitudes, beliefs and behaviors among PW, a rare population group. There is a critical need to provide these data as quickly as possible for actionable modifications in public health communications and partnership efforts to promote vaccination. Using a non-probability Internet panel sample is deemed fit for this purpose and application. However, it is not always appropriate to make statements regarding the generalizability of estimates based on non-probability samples, as it would be if the estimates were based on probability samples. Our exploratory research investigates and develops procedures to ex post facto match the 2013 non-probability Internet panel sample of PW with a probability-based sample derived from the 2013 National Health Interview Survey (NHIS). The objective is to improve the ability to assess the precision and bias of the vaccination coverage estimates from this non-probability sample of PW. The steps used to identify and classify NHIS PW data using the same or similar definitions as the PW Internet panel data for information reported for a similar time window are presented. Vaccination coverage estimates are calculated using weights for both samples raked to the identical population control totals. Variables are identified that contribute to observed differences between internet panel and NHIS estimates. We demonstrate matching a sample from the non-probability panel to a probability sample of PW using a set of variables that will make the non-probability sample similar to the probability sample. A likely precision of vaccination coverage using this matched sample is calculated.

Weighting Web Panel Data at the Community Level

Ronaldo Iachan, *ICF International*

John Boyle, *ICF International*

James Dayton, *ICF International*

Lew Berman, *ICF International*

While non-probability web panels can generate an adequate sample size for many communities, the sample must yield estimates that can be considered accurate and representative. This paper looks at an internet panel survey designed to provide representative data for three local areas—one urban county, one 5-county city and one city. The ongoing survey weighting process includes post-stratification adjustments to the American Community Survey (ACS) population data for each local area. Using an ACS variable on Internet access, we investigate whether internet panels provide accurate data about the subpopulation of persons with internet access, and how these compare with other survey data for the same subpopulation. The weighting process will be performed in stages that allow for the comparisons of the pre- and

AAPOR 70th Annual Conference

post-adjustment data to similar estimates available from the ACS and the BRFSS for the same locality. One initial comparison with ACS data for the local area capitalizes on an ACS variable on Internet access, a variable that seems most critical for the adjustment of Internet panel data. This variable defines two subsets, people with and without Internet access. We will assess different weighting approaches that consider this bi-section. We will also discuss the second stage of weighting that goes beyond ACS-type demographics to adjust for BRFSS-type variables related to lifestyle and health risks, also included in our panel survey questionnaire. Example variables are related to health insurance status, health status, smoking and cancer screening. We compare weighted estimates for other health outcomes before and after this additional step of weighting.

Computer-Administered Self-Interviewing (CASI): Impact of Setting, Interviewer and Voice Characteristics

Can Interviewer Behaviors During ACASI Affect Data Quality?

Emilia Peytcheva, *RTI International*

Brady West, *University of Michigan*

Many national face-to-face surveys use audio computer-assisted self-interviewing (ACASI) to collect sensitive information from survey respondents. ACASI is used to mitigate concerns about education and literacy, and ensure privacy when respondents are providing answers to sensitive questions. However, human interviewers remain present during ACASI to introduce the technology, and potentially provide assistance. Interviewers may vary in their ability to establish a private environment and in the level of assistance that they provide, but no research to date has examined this possibility in any level of detail. This study examines the effects of certain interviewer behaviors during ACASI, ascertained from post-survey interviewer observations, on the response distributions for selected sensitive survey items in the National Survey of Family Growth (NSFG). We examine the amount of variability among NSFG interviewers in selected behaviors during ACASI administration, such as sitting arrangement relative to the respondent, ability to see the respondent's screen, and provision of assistance. We analyze these behaviors from 22,682 NSFG interviews, completed by 108 interviewers. The results show that the prevalence of these behaviors varies significantly among NSFG interviewers. In addition, we find that these behaviors do in fact affect the response distributions of some sensitive NSFG items, even after controlling for socio-demographic characteristics of the respondents. This study provides motivation for future randomized experiments that explore the effects of these interviewer behaviors further, and has the potential to improve the training of interviewers on what behaviors to avoid for maximizing the quality of data collected using ACASI.

Effects of ACASI Voice Choice and Voice Persona on Reports to Questions About Sensitive Behaviors Among Young Adults

Kerryann DiLoreto, *University of Wisconsin Survey Center*

Jennifer Dykema, *University of Wisconsin Survey Center*

Karen Jaques, *University of Wisconsin Survey Center*

Nadia Assad, *University of Wisconsin Survey Center*

Audio computer-assisted self-interviewing (ACASI) – in which respondents answer pre-recorded questions during an in-person interview – is the preferred method for administering sensitive questions in face-to-face interviews because it often yields higher reports of sensitive behaviors

(Tourangeau and Smith 1996; Turner, Ku, et al. 1998). In addition, the inclusion of audio may overcome barriers to reading among respondents' with low literacy levels. However, research raises doubts about whether respondents actually use the audio and whether the inclusion of audio has any effect on survey responses. In order to increase the likelihood that respondents would refrain from turning the audio off, we implemented an experiment in which respondents were randomly assigned to hear one of three types of prerecorded voices versus being presented with the option of selecting from among the three voices. Voices were selected to represent different personas, including an empathetic-sounding voice, a professional-sounding voice, and a synthetic (text-to-speech) voice. We hypothesized that providing respondents with the choice of voice would mitigate the tedium of listening to the audio and emphasize the importance of the audio component. We examine the effects of voice choice and the type of voice listened to on levels of reporting about sensitive behaviors, item nonresponse, and the proportion of audio listened to. We also examine the effect of literacy on respondents' propensity to turn the audio off. Data are from the California Youth Transitions to Adulthood Study in which 727 foster care youth (aged 17 in 2013) participated in an in-person interview about their experiences while wards of the state (RR=95.3%). Questions asked about a variety of sensitive topics including delinquency, violence, and sexual behaviors. Our analysis adds to the small body of research on voice effects in surveys using ACASI, with important findings on measuring sensitive behaviors, particularly among young adults.

Race-of-Interviewer Effect in the Computer-Assisted Self-Interview Module in a Face-To-Face Survey

Mingnan Liu, *SurveyMonkey*

Yichen Wang, *NERA Economic Consulting*

This article explores how interviewer race influences responses to race-related questions in a Computer-Assisted Self-Interview (CASI) module of a face-to-face interview. Utilizing the 2012 American National Election Studies (ANES), this study examines two sets of race-related questions. One set of questions asks how hardworking whites, blacks, and Hispanics are. The other set asks respondents to rate their feelings toward whites, blacks, and Hispanics on feeling thermometers. The racial priming effect and social desirability theories predict that, when the interviewer's race matches that of the rating object, respondents are more likely to express positive opinions of that race on both the hardworking and feeling thermometer questions, in comparison to when the interviewer's race does not match that of the rating object. Our results provide partial support for the race-of-interviewer effects. For the "hardworking" questions, we find race effects when the interviewer is black or Hispanic but not when he or she is white. For the feeling thermometer questions, we find significant race effects with white and Hispanic interviewers but no such effect with black interviewers.

Comparison of Text-To-Speech with Human Voice Recordings on Comprehension of Survey Questions in Audio Computer-Assisted Self-Interviewing

Emily McFarlane Geisen, *RTI International*

Gretchen McHenry, *RTI International*

Patty LeBaron, *RTI International*

Gil Rodriguez, *RTI International*

Grace O'Neill, *SAMHSA*

Peggy Barker, *SAMHSA*

Dicy Painter, *SAMHSA*

Joel Kennet, *SAMHSA*

Advances in Text-to-Speech (TTS) technology have enabled realistic, accurate, and clear-sounding voices that can be easily programmed and customized for use in audio computer-assisted self-interviewing (ACASI). However, the comprehension of survey questions using TTS-generated voices compared with human voice recordings has not been evaluated with the individuals who are most likely to rely on the audio component when completing the ACASI portion of an interview. These individuals are likely to be the youngest and oldest respondents (i.e. those aged 17 or younger and 65 or older), as well as respondents with low levels of literacy and non-native English speakers. This paper presents results of cognitive testing completed with these types of participants on the National Survey of Drug Use and Health (NSDUH) to compare two TTS voices with the standard, human voice used in the current NSDUH. Thirty-six cognitive interviews were conducted in both English and Spanish. All participants were presented audio recordings of survey questions using three voices: human voice, TTS at a moderate speed, and TTS at a slow speed. Participants only heard the questions and were not able to see the questions. After hearing each survey question, participants were asked comprehension probes to assess how well they understood the words or concepts presented in the question. For example, participants were asked to repeat questions or explain the meaning in their own words. All participants received the same questions and the same probes in the same order, but the order of the voices presented was varied. Although both English- and Spanish-speaking participants seemed to prefer the human voice slightly, most participants thought the TTS voice was pleasant and understandable. More importantly, we found no differences in comprehension ratings between the three voices for either English- or Spanish-speaking participants, indicating that TTS is a viable option for ACASI.

Building and Evaluating Likely Voter Models

The Choice is Yours: Comparing Alternative Likely Voter Models within Probability and Non-Probability Samples

Robert Benford, *GfK*

Randall K. Thomas, *GfK*

Jennifer Agiesta, *The Associated Press*

Emily Swanson, *The Associated Press*

Likely voter models often improve election predictions – for both voter turnout and vote choice. Successful modeling typically combines several measures to estimate registered voters, voter turnout, and vote outcome. A number of factors have made likely voter modeling more difficult, including the broader utilization of early voting and changes in sampling and data collection

AAPOR 70th Annual Conference

mode. In October 2013 the AP-GfK Poll moved from dual-frame RDD telephone surveys to an online protocol using KnowledgePanel®, which is the largest U.S. probability-based online panel, enabling rapid and detailed national polling. Though KnowledgePanel is sufficient for national projections, one of the key interests is prediction of voting outcomes by states where KnowledgePanel can fall short. As such, GfK and The Associated Press (AP) have examined how larger demographically balanced non-probability (opt-in) samples could supplement probability-based (KnowledgePanel) through a calibration methodology. To study this, we selected two states with diverse populations - one in the Midwest often favoring Democrats (Illinois) and one in the South often favoring Republicans (Georgia). Each state had both Senatorial and Gubernatorial races on the line. In each state two parallel surveys with about 800 KnowledgePanel and 1,600 opt-in respondents were administered immediately prior to the elections. Subsequently, about 70% of pre-election respondents were surveyed again post-election. Respondents in each sample were randomly assigned to one of two alternative sets of likely voter items - either the AP-GfK Poll's standard likely voter set or an alternate set driven mainly by stated intention to vote. We report estimates of registered voters, turnout, and election results by likely voter model, how that model can be optimized, and comparisons of estimates separately from KnowledgePanel and opt-in samples. While both models predicted well, the revised model used fewer variables. In addition, calibrating opt-in samples to probability samples can improve the utility of opt-in samples.

Voter Turnout Sensitivity Analysis: Towards a More Parsimonious Combinatorial Likely Voter Model

Joseph Chris Jackson, *Ipsos Public Affairs*

Julia Clark, *Ipsos Public Affairs*

Clifford Young, *Ipsos Public Affairs*

Typically 'Likely Voter Models' include a 5 or 6 item summated index from which "cut points" are defined, which typically correspond to the historical turnout rates for a given type of election (Presidential or Midterm). Recent work has explored the challenges of this model, specifically the 'coarse' nature of the index, which does not allow for as much precision as needed to approximate expected turnouts. The problem is especially acute in Primary or low-turnout Midterm elections like 2014, where only 25% - 35% of the population cast a vote. The problem has been in part addressed via the use of logistic regression models that allow for more discrimination between intervals. However, the need for a simple question battery allowing for sufficient granularity at low-turnout elections persists. Our paper will undertake a sensitivity analysis for a broad range of turnout cutpoints from over 30,000 interviews conducted in the run-up to the 2014 Midterm elections, for both generic congressional ballot and state-specific voting intention measures. We will re-analyze the data in order to assess both question sensitivity and the full Likely Voter battery, in an attempt to create a combinatorial model that is more parsimonious. To measure performance, we will employ the Average Absolute Difference between the survey estimates and election results.

The Effects and Effectiveness of Likely Voter Models in Pre-Election Surveys

David L. Vannette, *Stanford University*

Jon A. Krosnick, *Stanford University*

Many pre-election polls identify the preferences of people labeled "likely voters". One challenge in pre-election polling is the fact that often a substantially greater proportion of respondents report that they will vote than the proportion of the population who actually end up voting. Consequently, researchers wish to identify the subset of respondents who are truly likely voters,

AAPOR 70th Annual Conference

and different organizations and academic researchers use different approaches. Our study evaluates the effectiveness of a variety of different methods for identifying likely voters. We use data from the American National Election Studies (ANES) 2008 Time Series Study, which administered many measures that could be used to identify likely voters during pre-election interviews. Then, these same respondents were interviewed post-election and were asked whether they in fact voted. We attempt to identify the optimal combination of pre-election reports with which to effectively separate people who did vote from those who did not. We use the ANES vote intention question as a starting point and supplemented it with a variety of non-demographic and demographic measures that are known to predict voter turnout. We explore several varieties of two basic, but methodologically distinct approaches: (1) dividing respondents into two categories: voters and non-voters via a discriminant function approach, and (2) assigning a probability of turnout to each respondent and then comparing the effectiveness of various cut points along the probability dimension for separating voters from non-voters. We assess model accuracy using post-election reports of turnout and estimates of candidate vote share along with external benchmarks of the demographic composition of the 2008 electorate. The result is evidence pointing to the methods that can be most effective in identifying which survey respondents will vote on Election Day. This evidence will be of value to all researchers who conduct pre-election polls or who interpret such data.

Using Voter File Validation to Improve Likely Voter Models: The Case of 2014

Ruth Igielnik, *Pew Research Center*

Scott Keeter, *Pew Research Center*

Michael Dimock, *Pew Research Center*

Jocelyn Kiley, *Pew Research Center*

Kenneth M. Goldstein, *University of San Francisco*

As has been amply reported, there was a significant Democratic bias in the pre-election polling leading up to the 2014 midterm elections. Polls were biased, on average, about four points in a Democratic direction. The fact that voter turnout in 2014 was down significantly from 2010, and that Democratic candidates suffered losses up and down tickets suggests that inaccurate forecasts about the turnout of different key groups, and thus the composition of the likely electorate, were to blame for the bias. A definitive conclusion about the source of the errors will require data from a range of pollsters, but we address one aspect of the issue – the predictive accuracy of a variety of “likely voter” measures – with the Pew Research Center’s American Trends Panel, a nationally representative panel of more than 3,000 adults surveyed before and after the election about their preferences in U.S. House elections. More than 80% of the panelists were matched to a national voter file, allowing us to validate whether or not they voted in the election. Knowing this, we can look at numerous pre-election measures to assess their ability to accurately predict voting. These include (1) measures of campaign interest and engagement, plus measures of general political engagement; (2) measures of self-reported intention to vote; (3) measures of self-reported past voting behavior, plus past voting from the voter file; (4) ideological consistency across 10 questions about issues and values; (5) intensity of partisan sentiment; (6) proprietary voter file propensity to vote. Unlike other previous likely voter studies, this project has the benefit of an extremely large sample size, a match rate that exceeds the typical voter validation effort, a panel design, and a variety of measures that can be used to simulate different styles of turnout screens with both RDD and list samples.

AAPOR 70th Annual Conference

Estimating the 2014 National House Vote: What Can Be Learned

Lydia Saad, *Gallup*

Frank Newport, *Gallup*

Jeffrey M. Jones, *Gallup*

Stephanie Kafka, *Gallup*

This paper focuses on the results of 2014 estimates of the national House vote (the "generic congressional ballot") as a mechanism for understanding the ways in which surveys can estimate real world behavior, a validation function that is one of pre-election survey's important functions. Final estimates of the House vote from various survey organizations clustered in two groups, some relatively close to the final aggregated House vote, and others underrepresenting the Republican share of the vote. The paper reviews the methods and methodological assumptions which may have been responsible for the differences, and include analysis of experimental research conducted by Gallup in 2014, using both traditional RDD and a web-based probability panel. The results shed light on the impact of fundamental, underlying sampling methodologies, as well as the impact of likely voter modeling, on the accuracy of the 2014 estimates of Congressional voting.

Freedom and Other Hot Topics in Public Opinion

Global Approval of U.S., EU and Russia's Leadership on the Brink of a New Cold War

Julie Ray, *Gallup*

Neli Esipova, *Gallup*

Anita Pugliese, *Gallup*

As East-West tensions over the Ukraine crisis in 2014 push the world to the brink of a new Cold War, which "side" the populations of different countries are on – whether they support the leadership of the U.S., European Union, or Russia – likely will play a role in their foreign policy going forward. For the past seven years, Gallup World Polls in 130+ countries have shown Russia's leadership earning the lowest approval ratings in the world compared with other world powers. A median of 24% worldwide approved of Russia's leadership in 2013, but it continued to maintain a solid base of support among its neighbors (formerly Soviet Union countries). Although the U.S. has seen its approval ratings ebb and flow, it has usually received higher ratings than other world powers. In 2013, its median approval stood at 46%. Approval ratings for the leadership of the European Union have typically fallen somewhere in between Russia's and that of the U.S. This paper will examine what effect the standoff between the U.S. and Europe and Russia in 2014 might have had on public opinion worldwide, based on World Poll surveys in 130+ countries. It will also discuss what the sphere of influence of the leadership of the U.S., EU, and Russia looks like today in the increasingly tense global atmosphere.

A Cross-Cultural Look at Religious Tolerance

Caitlin E. Deal, *University of Nebraska, Lincoln; Survey Research and Methodology*

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

This study examines the degree to which religious tolerance differs between countries with different levels of religiosity. Drawing on previous research examining religious tolerance in the

AAPOR 70th Annual Conference

United States and the United Kingdom (Deal, Countryman and McCutcheon 2014), we expand the investigation to include religious tolerance in Germany, a country that has had a very different history regarding religion over the past several generations. The U.S. population is much more religious than the UK, with 67.7 percent of U.S. respondents stating that religion is an important part of their daily life, compared to 30.7 percent of those in the UK (GWP, 2008). Previous research, using latent class analysis, shows that it is not the religiosity of the country that predicts religious tolerance, but how important one's religion (what the religion is does not matter) is to them in their daily life (Deal et al, 2014). We examine Shils' theory of center and periphery to investigate the importance of identification with the in-group (in this case, religion) as a source of intolerance. This research examines religious tolerance in Germany to see how it compares to the findings in the U.S. and the UK. Specifically, we examine the separation between the East and West, where Eastern Germany was far more secularized than Western Germany (Pollack & Pickel, 2007). Preliminary analyses of data collected by the Gallup World Poll, which uses a probability based sample, from October of 2008 show that 38.4 percent of Germany respondents report that religion is an important part of their daily lives, and 61.6 percent report that it is not. These responses suggest that religion in Germany is intermediate to the UK and the US; we will examine how this middle ground affects religious tolerance.

Afghanistan: After the Presidential Transition

Matthew Warshaw, *Afghan Center for Socio-economic and Opinion Research*

Gary Langer, *Langer Research Associates*

Stephen Hornbeck, *D3 Systems Inc.*

Amanda Bajkowski, *D3 Systems Inc.*

Christopher Weiss, *Langer Research Associates*

Another disputed election in Afghanistan produced a compromise result in 2014, with Ashraf Ghani sworn in as president and his challenger, Abdullah Abdullah, awarded a newly created position as CEO of the Afghan government. How do Afghans view this outcome, the departure of most Western forces and the position of their country as it embarks on its "transformation decade"? ACSOR Surveys, D3 Systems, and Langer Research Associates have collaborated on a national poll in Afghanistan, fielded in late October and early November 2014, to examine these and related attitudes. Our goal is to continue to enrich public understanding of Afghanistan, as well as to share insights into the methodological and operational challenges of conflict-zone research.

Ukraine and the West vs. Russia and the Rest: The Media Battle for Public Opinion

Neli Esipova, *Gallup*

Julie Ray, *Gallup*

Dato Tsabutashvili, *Gallup*

Gallup World Polls show the vast majority of people living in former Soviet republics were following the news about the situation in Ukraine and Crimea very closely during the conflict in 2014, with most believing the Russian state media's version of events over the news from private Russian media and the West. Majorities in many countries supported Crimea joining Russia, while just a few stood steadfastly against it. Only 7% in Georgia supported it, as did only 10% in Azerbaijan. In all countries, however, some percentage of the population believes what happened in Crimea could happen to them – regardless of whether their country shares a border with Russia. No one dismisses the possibility outright. This paper will analyze public

AAPOR 70th Annual Conference

opinion data from residents of 12 countries surveyed between May and October 2014, looking specifically at their attitudes toward the conflict, the role media likely played in forming these opinions, and how it changed the way they see the future of their relations with Russia, the U.S. and the European Union. This paper will also feature new data collected in Ukraine in fall 2014 and discuss some of the challenges Gallup faced in obtaining data during the conflict.

What Defines Democracy? Public and Elite Perceptions of Democratic Norms and Structures

Frank Louis Rusciano, *Rider University*

Josephine Boyle, *Rider University*

Michael Brogan, *Rider University*

The norms and structure of democratic procedures have occupied writers for centuries. Added to the questions about what constitutes a democracy are the variations that exist in different nations in the world. Elite measures as The Combined Index of Democracy (KID) and the combined index of democracy 3 dimensions (KID3D) measure the regime quality of currently 165 countries every two years over the period 1996 to 2012. Similarly, the CIRI measure named after its creators Cingarelli and Richards, provides information about government respect for a broad array of human rights in nearly every country in the world. Norris (2012) opens an important area of inquiry by comparing these elite measures with citizen perceptions of democracy across various nations using the World Values Survey (WVS). She discovers that there is a significant correlation between the elite perceptions of democracy and citizen perceptions of democracy, as measured on the national level using mean values from the WVS. This paper's analysis extends Norris's study, using an approach designed by Rusciano (2012, 2013) to compare elite and citizen perceptions of corruption. In these studies, citizen perceptions correlated with such factors as service provision and government power beyond the elite measures, to provide new insights into how to define corruption globally. This paper applies the same approach to perceptions of democracy, by asking what other factors beyond the elite measures explain the variance in citizen perceptions of democracy across various nations. It queries further whether these additional measures provide clues about what constitutes democracy on a global level. The paper concludes by speculating if we may begin to define a "world opinion" whether and how democracy might be defined across nations using the elite and public measures.

Panel: Training for Survey Research: Who We Need? Where We Train Them? And, How Are We Going to Fill the Jobs of the Future?

Panel: Training for Survey Research: Who We Need? Where We Train Them? And, How Are We Going to Fill the Jobs of the Future?

Graham Kalton, *Westat*

What would be the ideal person you would like to hire into a survey methodologist position? A diverse team of researchers in industry, government and academia shares their ideas about the skillsets they usually see in survey methodologists in their organizations, and outlines what they think an "ideal" survey methodologist entering the work force of a survey organization should be familiar with. We present the key competencies expected of survey methodologists in the statistical and social science components of a survey methodologist's training, as well as the key "soft" skills that are necessary to be successful in a typical survey methodologist position.

AAPOR 70th Annual Conference

We also look at the ways in which academic institutions can respond to these demands by offering new programs and new educational formats. Developments that led to this panel are based on an earlier panel on training for survey research held at Pew Research Center on June 11, 2014. Questions and issues that were raised at the 2014 panel have been critically assessed, reformulated, and further developed into the topics that are presented now.

Training Needs in Survey Research Methods: An Overview

Graham Kalton, *Westat*

The last quarter century has seen an explosion in research on survey methods in both the statistical and non-statistical aspects of the field, an explosion that has been accompanied by a corresponding growth in the survey methodology literature. The period has also seen important developments in the training opportunities for survey methodologists at universities and in professional organizations. Survey methodology now has a well-established standing as a discipline of specialization. Furthermore, survey methodology is in a state of transition for many reasons, including falling response rates and alternative modes of data collection. Much research and many new developments are to be expected in the next decade. This paper provides an overview discussion of the types of training programs needed to prepare those intending to enter the field of survey methodology and of the training opportunities needed to keep survey methodologists up-to-date on new developments over the course of their careers.

Training for the Modern Survey Statistician

Stanislav Kolenikov, *Abt SRBI*

This paper describes the set of skills that a modern survey statistician must possess to be effective in research, study design and survey analyses positions, be that in academia, industry or government. We assume that such a survey statistician is receiving his or her training in a “regular” statistics department or in one of the (less than a handful of) survey methodology programs. We outline the set of statistical and survey techniques that are typically expected to be found at the Master’s and at the Ph.D. level. We also outline the minimal set of statistical methods that any survey methodologist must be proficient in.

Social Science Survey Methodology Training: Understanding the Past and Addressing the Present to Shape Our Future

Scott Fricker, *Bureau of Labor Statistics*

Matthew Jans, *UCLA*

Mikelyn Meyers, *U.S. Census Bureau*

This paper addresses training issues for social science survey methodologists. We begin by reviewing the state of training in the field, and how it fits into a vision of survey methodology as a historically-interdisciplinary cum independent academic discipline. Focusing on social science aspects of survey methodology training, we review curricula of existing survey methodology programs, list training gaps as we see them, and outline an ideal program of study. We emphasize the importance of statistical training and cross-disciplinary theory exposure in the training of social science survey methodologists at both the Masters and Doctoral levels.

AAPOR 70th Annual Conference

On-the-Job: What to Expect and How to Succeed

Mikelyn Meyers, *U.S. Census Bureau*

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

Stanislav Kolenikov, *Abt SRBI*

Matthew Jans, *UCLA*

The work environments of survey methodologists are varied and their roles are diverse. Across commercial, government and the academic sectors, survey methodologists are usually involved in conducting surveys for practical use by stakeholders who need the data for decisions unrelated to “survey methodology” per se. Besides the “academic” skills discussed in the previous two papers, survey methodologists invariably find themselves in collaborative positions. Both substantive and survey expertise are needed to design, execute and produce estimates. Survey teams with representation of all the knowledge and skills related to the survey topic and objectives are typical and collaborate from concept to final product. Employers need staff with a good understanding of the survey lifecycle and sources of survey error. Good communication skills are highly valued. We discuss this and other non-technical skills required for a survey methodologist to succeed in their position. The field of survey methodology is evolving with many new challenges to the science of accurately measuring attitudes and behaviors. There is both a need for and many opportunities to obtain continuous learning throughout the career of a survey methodologist. We outline some of the existing opportunities to continually learn in the work environment, particularly for those early in their career.

Survey Informatics: The Future of Survey Methodology and Survey Statistics Training in the Academy?

Allan McCutcheon, *University of Nebraska, Lincoln*

Jill Dever, *RTI International*

Declining response rates and the increasing costs of traditional means of data collection have led to a dramatic surge in Internet surveys. The near certain move to the Internet for the first wave of data collection for the 2020 Census, coupled with diminished governmental budgets for data collection, are likely to accelerate the already rapid move of survey data collection to the Internet. These changes hold a great deal of promise for the survey research community – the relative ease with which we can collect paradata, passive data collection, the introduction of ‘personal assistants’ (e.g., Siri, Cortana) as potential survey interviewers, all point to new areas in which future developments are already becoming realities. Importantly, these changes will also require a rethinking of academic curricula aimed at training the next generation of survey methodologists. I will explore the emerging field of survey informatics in which new, and more prominent, roles will be played by computer science, statistical and mathematical modeling, complementing the traditional role of cognitive and social psychology. These developments will require greater collaboration with professionals in the field, and are also likely to confront challenges and barriers to innovation in the academy.

Weighting and Imputation

Multiple Imputation for Complex Surveys: An Overview of the State of the Art

Joseph L. Schafer, *U.S. Census Bureau*

Multiple imputation (MI) was proposed by Donald Rubin in 1977 as a general paradigm for handling nonresponse in surveys, censuses and administrative databases. Over the last decade, it has gained widespread acceptance in many fields of quantitative research, and nearly every major statistical software package now performs MI in some fashion. In some cases, MI has been applied to complex surveys, but in other cases survey practitioners still find MI too daunting to implement. In this presentation, I review the history of MI, describe the statistical and nonstatistical challenges it poses in the survey world, and look ahead to future developments in this area. Statistical issues include: mismatch between the imputation model and subsequent analyses; the challenges of jointly imputing large numbers of survey variables of mixed types; the need to preserve complex interactions; the need to reflect important features of the sample design; hierarchical or multilevel structures with missing values at multiple levels; preservation of skip patterns and logical constraints. Nonstatistical issues include: the availability and utility of software; perceived and actual difficulty of carrying out MI in a production environment; perceived and actual difficulty of explaining MI to data users and clients; and organizational culture and priorities.

The Influence of Attrition Weights on the Evaluation of Measurement Reactivity in an Intensive Longitudinal Study

Jamie Griffin, *University of Michigan*

Megan E. Patrick, *University of Michigan*

Measurement burst designs combine short-term and long-term longitudinal methods (e.g., a daily diary study recurring annually) and, thus, permit the study of within-person and between-person associations over time (Nesselrode 1991; Sliwinski 2008). Despite the important advantages of these designs, attrition and measurement reactivity (panel conditioning) can affect data quality (Iida et al., 2012; Warren and Halpern-Manners, 2012). Disentangling measurement reactivity from attrition bias, however, can be difficult (Kalton et al. 1989). Using an experimental design, we evaluated measurement reactivity in a study of young adult substance use. In the spring of 2012, high school seniors (N=300) from three Midwestern schools completed a baseline survey. Two-thirds of participants were randomly assigned to an intensive measurement (IM) group and the remaining third were randomly assigned to a control group. Young adults in the IM group received a 30-minute follow-up survey followed by 14 daily surveys four, eight, and twelve months after baseline (Waves 1, 2, and 3, respectively); young adults in the control group received only the Wave 3 follow-up survey. Response rates to the Wave 3 survey were 34.2% and 40.2% for the IM and control groups, respectively. First, to address attrition, we developed four sets of attrition weights based on response propensity models incorporating either (1) demographic information, (2) baseline substance use behaviors, (3) paradata, or (4) all of (1) through (3). Next, to evaluate measurement reactivity, we compared the weighted Wave 3 responses of the IM and control groups. When attrition weights were based on demographic information or baseline substance use behaviors, measurement reactivity was detected. Conversely, when attrition weights were based on paradata or all auxiliary variables, there was no evidence of reactivity. These findings highlight the importance of properly specifying response propensity models and the need for more research about best methods for doing so.

Examining Best Practices for Sampling and Weighting Dual-Frame Surveys

Elizabeth Kantor, *Rutgers School of Arts and Sciences Honors Program*

In recent years, the task of survey methodologists has been complicated by the necessity of incorporating cell phone respondents into survey samples due to the demographic and attitudinal distinctiveness and the geographic mobility of the increasing cell-only population. In addition to changes in sample composition, long-standing statistical techniques to reduce bias (i.e. systematic error), such as weighting, must be altered in order to properly incorporate cell phone respondents. In making choices about how to incorporate these methodological concerns, survey researchers must balance considerations of statistical variance, financial costs, and the extent to which their methods reduce bias. This research seeks to determine the best practices for sampling and weighting surveys under these considerations, using the example of the Rutgers-Eagleton Poll, a New Jersey university-based statewide public opinion poll that incorporates both cell phones and landlines in a dual frame, random digit dial design. I confirmed the demographic distinctiveness and higher level of geographic mobility of the cell-only respondents of the Rutgers-Eagleton Poll, as predicted by the literature. Then, four proportions of cell users in the sample were tested, and I found that a sample composed of 50 to 60 percent cell phones best replicated the cell-only estimates of the population from the Centers for Disease Control, though these designs would come at a higher financial cost. I applied sixteen combinations of weighting and trimming methods to a recent Rutgers-Eagleton Poll, and analyzed their effects on the variance and on bias reduction. Using criteria developed from this analysis, I recommended three possible weighting combinations, which utilized a base weight, and a percentile cap, and yielded a cell-only estimate within one percent of CDC estimates.

Methodological Briefs: Online Surveying and Recruiting - Recent Developments and Technological Innovations

How Fast Can I Get Survey Results? Assessing Demographics Differences in Cumulative Daily Response Rates to Web Panel Surveys

Kirti N. Kanitkar, *Gallup Inc*

Jenny Marlar, *Gallup Inc.*

Web –based panels are an increasingly popular mode of survey research. However, not enough is known about response rates among different demographic populations through the duration of the survey field period. Such differences in early versus late responders can have implications for length of survey period and sample size estimations. The present study will address three research questions. First, do internet panel response rates differ by age, education and race/ethnicity? Second, are there differences in cumulative daily response rates by age, education and race/ethnicity? Third, does a longer field period capture harder to reach respondents that do not respond early in the field period, or do they tend not to respond, regardless of time? The study uses data from the Gallup Panel, one of the few probability based mixed mode Panels which maintains demographic profiles of all its members. This allows for examination of daily cumulative response rates by demographic profiles. All members of the Gallup Panel are invited to complete surveys on a regular basis on topics ranging from Media Usage to Current Affairs to Well-Being. Using three such studies from 2014, this study examines daily cumulative response rates across different key demographics. Initial results indicate that younger, less educated and minority respondents have lower overall survey

AAPOR 70th Annual Conference

response rates in a probability based Web Panel. There were also differences by demographics in cumulative daily response rates. The rate at which the cumulative response rate increased from Day 1 to Day 7 and then to Day 14 was different across demographic groups, indicating early responders are demographically different from late responders. This paper also discusses implications of differences in fast vs slow responders on survey estimates as well as study design decisions such as study field periods and sampling techniques.

A Systematic Approach to Usability Evaluation of Web Survey

Lin Wang, *U.S. Census Bureau*

Temika Holland, *U.S. Census Bureau*

Marylisa Gareau, *U.S. Census Bureau*

The Internet has now become a common mode for survey data collection. However, usability problems with Web survey instruments can contribute to both measurement error (e.g., error in data entry) and nonresponse error (e.g., survey drop out). Usability, in this context, refers to the extent to which a respondent can self-administer the Web survey effectively, efficiently, and satisfactorily. In other words, a Web survey instrument with good usability would allow the respondent to complete the survey without errors, in reasonable time, and with a positive attitude towards the survey experience. Although well-designed usability evaluations for Web survey instruments can help identify and address usability problems and consequently reduce measurement and nonresponse error, optimally designing such evaluation requires more guidance than has generally been available. Common Industry Format (CIF) is an international standard (ISO/IEC 25062:2006) for reporting usability evaluation. CIF gathered industry-best-practices in three major areas: evaluation design, data collection and analysis. Drawing upon the CIF guidance, we propose a systematic approach to designing and conducting usability evaluations. This approach involves the specification of (1) evaluation objectives, (2) respondent performance metrics, (3) target respondent population, (4) evaluation participants sampling and recruiting methods, (5) data collection methods, (6) data analysis methods, (7) reporting of results. In this presentation, we will first introduce the systematic approach, followed by using the approach to review the usability evaluation of a national survey to demonstrate the approach's usefulness. We will particularly examine the measurement of survey administration effectiveness, efficiency, and satisfaction. Methodological and operational challenges in the application of this systematic approach will be discussed accordingly.

Attrition in a Probability-Based Mixed-Mode Panel: Does Survey Mode Matter?

Bella Struminskaya, *GESIS - Leibniz Institute for the Social Sciences*

Michael Bosnjak, *GESIS - Leibniz Institute for the Social Sciences*

For longitudinal surveys, panel attrition is a major concern as it reduces the sample size and may result in biased estimates. Mixed-mode panels face an additional complication if attrition differs between the modes. In order to maintain panel representativeness, methods that correct for attrition need to account for such mode effect. However, few published studies have focused on attrition in mixed-mode panels yet. In this paper, we seek to answer the question: Does panel attrition differ by the survey mode? We use data from the GESIS Panel, probability-based mixed-mode general population panel, which has online and mail components. For the first four waves of the panel, attrition rate for the offline participants is twice as high as for the online participants (roughly 6% vs. 12%, N=4886). Furthermore, overall panel attrition is selective on certain demographic and survey experience variables. We explore whether demographic and non-demographic correlates of panel attrition differ between the online and the mail mode and what consequences it has for the representativeness of the panel as a whole. We conclude with

AAPOR 70th Annual Conference

the discussion of ways to address differential attrition in mixed-mode panel designs. This paper will be useful for existing panels that employ additional (e.g., online) component to enhance contact and reduce costs and for panels that start as mixed-mode panels.

What Day of Week Would You Like to Answer Our Survey? A Large-Scale Randomized Experiment

Maria Andreasson, *University of Gothenburg*

Conventional knowledge states that most respondents prefer to answer online surveys on weekdays, rather than on weekends. Hence, it is also common practice to preferably send out online survey invitations, via e-mail for example, between Monday and Thursday. However, little evidence of this being an advantage can be found in previous research. In order to gather more reliable evidence on when to dispatch online surveys in order to maximize response rates we conducted a large-scale randomized experiments where 1,600 respondents belonging to the Citizen Panel at the University of Gothenburg were randomly assigned to each day of the week (total n=11200). Further, we included questions in the survey about which days of the week they preferred to answer our surveys. The findings indicate that initially Fridays and Saturdays, closely followed by Sundays, are the worst days to dispatch an online survey, all in accordance with conventional knowledge. On field day one, these days have significantly lower response rates than other days. However, after 3-4 days of data collection, the significant differences between dispatch days disappear, even without any follow-up reminder. This means that survey researchers do not need to consider which day of the week to dispatch a survey, unless they really want quick answers and an extremely short period of field work. Further, and perhaps more peculiar, respondents are more likely to say that they prefer to answer surveys on the day of the week when they actually received our invitation e-mail in the randomized experiment. Once they have tried Saturday, it does not seem so frightening anymore.

Boosting Response Rates in Online Longitudinal Studies: A Dose of Funny

Jessica M. Rath, *Legacy*

Andrea C. Villanti, *Legacy*

Valerie Williams, *Legacy*

Molly Green, *Legacy*

Paul Mowery, *Biostatistics Inc*

Donna M. Vallone, *Legacy*

Background: Although email invitations are a popular mechanism for soliciting participation in online research, email tone has not been examined as a factor influencing survey response rates. The purpose of this study was to determine if tone of email reminders influenced retention of young adult participants in a longitudinal study. Methods: Data were collected in August 2013 using Wave 5 of the Legacy Young Adult Cohort, a nationally representative cohort of 18-34 year olds. 5,190 participants eligible for follow up received emails containing a survey link for completion. The 2,963 non-responders were randomly assigned to one of three groups 1) usual/control email (N=732) 2) humorous email (N=1089) or 3) email with statistics from the study (N=1142). Logistic regression was used to generate odds ratios for the probability of completing the survey. Results: The experimental emails (humorous or statistics) increased response rates compared to the control email. After controlling for covariates, the odds of responding to the survey for participants who received the humorous email were 24% greater than for participants who received the control email (AOR=1.24, p=0.047). Regardless of study group, predictors of responding to the email invitation included: entering at earlier study waves vs wave 4, not smoking, living in smaller (< 5 people) vs larger households and being White vs

AAPOR 70th Annual Conference

Hispanic. Conclusions: Since online surveys consistently suffer from lower response rates relative to other survey modes, exploring factors that may affect online survey response rates is critical. Using tone to tailor reminder emails is one way to boost response rates in online surveys. Using humor in the email solicitation is a cheap, easy way to boost retention in a longitudinal study of young adults.

Comparing Social Media and Traditional Recruitment Methods: Which is Most Effective?

Jennifer Anderson, *Westat*

Jocelyn Newsome, *Westat*

Kerry Levin, *Westat*

April Oh, *National Cancer Institute*

Traditionally, researchers have relied on recruitment methods such as flyers, newspaper ads and Craigslist, to recruit participants for face-to-face qualitative studies. While historically effective, traditional methods are susceptible to attracting “professional” respondents. These approaches also rely on casting a wide net, thus making it difficult to target specific populations. With the rapid growth of social media, social science researchers increasingly turn to social media platforms as an alternative method for recruiting respondents for study participation (Hill et al. 2014). Social media recruitment can include methods such as Facebook ads, Google AdWords, LinkedIn ads, and outreach to bloggers and forums. Social media potentially offers several features not possible using traditional methods, including the ease of reaching respondents who may not otherwise be reached using traditional methods, the ability to target recruitment efforts to specific populations of interest (e.g., Millennials), and the ability to quickly adapt strategies based on preliminary outcomes and recruiting goals (Fenner et al. 2012; Hill et al. 2014). Furthermore, some studies suggest that social media strategies may provide cost savings (Chu & Snider 2013; Graham et al. 2012; Ramo et al. 2010). To date, however, there has been little research that compares the advantages and disadvantages of traditional and social media recruiting methods for face-to-face studies (Head et al. 2012). This paper will outline the results of a study that compared social media and traditional recruiting results for eight focus groups about health with women ages 20-50. During the six-week recruitment period, we tracked the success and costs of each method. We will discuss the advantages and disadvantages of each method in terms of costs per complete, ability to successfully recruit target populations, relative diversity of the pool, and show rates. We will share lessons learned and recommendations for optimizing recruitment using both methods.

Mobile Effects in Panel Surveys

Effects of Mobile versus PC Web on Survey Response Quality: a Crossover Experiment in a Probability Web Panel

Christopher Antoun, *University of Michigan*

Respondents are increasingly taking Web surveys on their smartphones instead of their personal computers (PCs). This change has received much attention, but the impact of completion device on response quality is still unclear. The study reported here compares data quality in mobile Web and PC Web. To reduce selection bias, I carried out a randomized two period, crossover design using the Longitudinal Internet Studies for the Social Sciences (LISS) probability panel. Participants (n = 1390) were invited to take both the mobile Web and PC Web

AAPOR 70th Annual Conference

surveys in sequence, with a month-long break in between surveys. Those who didn't have access to a smartphone were provided one for the experiment. Using multiple indicators of least-effort responding and socially desirable responding, I first test my hypotheses that respondents in mobile Web would satisfice more because of increased multitasking (and distractions) and disclose less because they are more likely to be in the presence of others compared to PC-Web respondents. I then assess response quality for several question formats that I expect to be sensitive to screen size (slider bar, spin wheel, and items with long lists of response options). I found that respondents in the mobile Web survey really were more mobile and more engaged with the other people and things around them compared to PC-Web respondents. Despite this, response quality – conscientious responding and disclosure of sensitive information – was equivalent between mobile and PC Web, even for those who were unfamiliar with smartphones. But for the items that were not effectively displayed on small screens, response quality was lower in mobile Web and this was consistent across subgroups. Overall, these results are encouraging for those already conducting carefully-designed mobile Web surveys; smartphones appear to have no negative effects on response quality if certain question formats are avoided.

The Changing Landscape of Technology and Its Effects on Online Survey Data Quality

Nicole Mitchell, *Survey Sampling International*

According to a report published in January of 2014 by Pew Research Center, 58% of U.S. adults own a smartphone. Since 2011, smartphone ownership has increased by 23 percentage points. More shockingly, the growth of tablet ownership increased from 3% in 2010 to 42% in January 2014. As ownership of these devices increase, it is important that we examine how the usage of these devices affects the quality of online data. To assess the impact these devices may have on the quality of online survey data, we conducted a study using SSI online sample consisting of 5000 U.S. respondents age 18+. Respondents were randomly assigned to one of 3 survey treatments: mobile unfriendly, mobile friendly and mobile optimized. The mobile unfriendly survey was a standard survey design, where regardless of device a traditional grid question was shown. The mobile friendly survey treatment was not designed for a mobile device, however the grid question was asked as multiple single punch questions across all devices. The mobile optimized survey was designed specifically for a mobile device, where the survey program recognized the device and optimized the survey for mobile respondents. The grid question was shown as multiple single punch questions on the mobile phone. However, on tablets, laptops and desktop the traditional grid format was shown. Each treatment group consisted of respondents from each device category: tablet, smart phone, laptop/desktop. We implemented various quality controls to examine whether satisficing behaviors differ between respondents who accessed our survey using a laptop/desktop, tablet, or smartphone. Furthermore, we examined whether optimizing the survey for the mobile phone impacted the quality of the data. Lastly, we investigated the consequences of excluding data received from mobile respondents.

The Effects of Adding a Mobile-Compatible Design to the American Life Panel

Alerk Amin, *RAND Corporation*

Vera Toepoel, *Utrecht University*

Peter Lugtig, *Utrecht University*

The American Life Panel (ALP) is a nationally representative, scientifically recruited online panel that began in 2006. Traditionally, respondents completed surveys using desktop web browsers.

AAPOR 70th Annual Conference

Computers and Internet access are provided to respondents who do not have Internet access. In September 2014, the survey design was updated to a responsive theme, allowing respondents to complete surveys using tablets or smartphones, in addition to desktop browsers. At the same time, a panel recruitment effort took place, adding over 800 new panel members. While most panel members saw the transition from the desktop-only design to a mobile-compatible design, the new panel members saw only the mobile-compatible design. This paper shows the mobile device adoption rate after the introduction of the new design, and compares the rates between the 2 groups of respondents, as well as other demographic variables. In particular, the change in behavior of the long-term panel members who saw the transition is examined. We discuss the implications of our findings for mobile-compatible web surveys.

Purposefully Mobile: Experimentally Assessing Device Effects in an Online Survey

Frances M. Barlas, *GfK*

Randall K. Thomas, *GfK*

Patricia Graham, *GfK*

The recent growth in the number of participants who take surveys on mobile devices can be seen as an opportunity to increase coverage and reduce non-response, but it is essential that online survey questionnaires be designed with mobile display in mind to improve data quality and minimize device-specific effects during data collection. Responsive design, where the survey display is adjusted based on the screen size and orientation, can help, but it cannot overcome all mobile survey design challenges. Evidence-based recommendations are needed to inform design decisions around survey question and response format. To develop such recommendations, we designed a study that randomly assigned respondents to one of two tracks, either a more mobile friendly track that incorporated shorter question stems and response scales or a less mobile friendly track that made no accommodations for mobile display. Both tracks were administered with a responsive design template. The study relied on GfK's KnowledgePanel®, the nation's largest online probability-based panel, and utilized panel members' profile data to identify and select participants who could take the survey on either a mobile or desktop/laptop device. Rather than rely on 'accidentally mobile' participants with their attendant self-selection bias, participants were randomly assigned to complete our survey on one of three devices (desktop/laptop, smartphone, or tablet). We had about 3,600 completed surveys with an average completion time of approximately 20 minutes. The experimental design allowed us to balance the samples to examine how device affects responses to various question and response formats, including single and multiple choice, horizontal and vertical display, open-ended numeric, sliders, and drop down formats over a variety of survey topics. We will discuss observed differences in responses that appear to occur because of device used and not due to sample differences and provide recommendations for designing online survey questionnaires to accommodate mobile respondents.

The Mobile Influence: How Mobile Participants Affect Survey Results

Frances M. Barlas, *GfK*

Randall K. Thomas, *GfK*

The number of surveys being conducted online has increased dramatically over the past 15 years. While most online surveys are designed for desktops/laptops, the proportion of respondents taking online surveys with a mobile device, either smartphones or tablets, is increasing, typically making up 20 to 30% of participants in general population studies today. Past research has found that mobile respondents differ from those who take surveys on

AAPOR 70th Annual Conference

desktops/laptops, and within mobile that smartphone respondents differ from those completing on a tablet. In general, smartphone participants are more likely to be male, younger, and minority respondents. Tablet participants are more likely to be female, middle-aged, and white. Typically, mobile respondents take longer to complete a survey and have a much higher suspend rate than those completing on desktop/laptop. In this study, we investigate differences in attitudes among these groups even after controlling for demographic differences. We used the Advertising Research Foundation's FOQ2 dataset, which had 57,104 completes from 17 different opt-in sample sources (each providing at least 3,000 completes) and fielded in January 2013. There were 51,641 who completed on a desktop/laptop, 2,276 on a tablet, and 2,089 on a smartphone. While this study replicated the demographic differences found previously for mobile respondents, we used these demographics (age, sex, region, race/ethnicity, education, and income) as covariates to equate the device groups and examine the attitudinal differences between the groups to identify any self-selection biases that were not accounted for by demographic factors. After controlling for demographics, there were no differences for issues like life satisfaction or need for cognition. However, there were significant differences for other issues; for example, tablet respondents were higher in technological orientation than smartphone respondents, who were higher than desktop/laptop respondents. We discuss the implications that increasing proportions of mobile participants will have on substantive survey results.

App vs. Web for Surveys of Smartphone Users

Kyley McGeeney, *Pew Research Center*

Ruth Igielnik, *Pew Research Center*

Smartphone ownership has steadily increased since 2011 and accordingly so has the use of mobile applications or "apps". The use of the internet for surveys has risen as well, with about a fifth of general population web surveys now being taken on mobile phones. A logical next step is to see whether we can marry the two trends, moving web surveys from a browser to an app on mobile phones. However this approach raises questions as to whether respondents would be willing to download an app and actually use it, and how the data collected via an app differs from data collected from a web browser survey on mobile. To answer these questions the Pew Research Center conducted an experiment that randomly assigned 2,146 members of a survey panel to receive a series of short surveys either on a mobile app or on a mobile browser. Of those invited, 91% have agreed to participate. The sample source for this experiment was the Pew Research Center's American Trends Panel, a nationally representative probability-based panel. This experiment used a signal-contingent experience sampling method to survey respondents twice a day for seven days about their smartphone use. Panelists were signaled via text message and email for both the mobile browser and app groups, as well as push notification for the app group. The study is in the field at this time and will be completed in November. Willingness to download the app will be examined as will response rate, non-response bias, and differences in substantive responses by treatment group. In addition to tabular comparisons of cooperating and non-cooperating groups, a multivariate analysis of the factors predicting consent and cooperation will be conducted.

Saturday, May 16
1:00 p.m. – 2:00 p.m.
Poster Session 3

Influences on Response Latency in a Web Survey

Benjamin Phillips, *Abt SRBI*
Stanislav Kolenikov, *Abt SRBI*
Elaine Howard Ecklund, *Rice University*
Allison Ackermann, *Abt SRBI*
Ho Won Cheng, *Abt SRBI*

Response latency to web surveys is of considerable interest. Time from the stimulus (here, displaying a question) to the response (here, recording the answer) is used to identify potentially problematic respondents (those with low latency) and items (those with high latency). Such uses are, however, typically narrowly constructed. We analyze a wide variety of factors using a rich dataset to develop a deeper understanding of the drivers of response latency in web surveys. The Rice University Religion and Science in International Context (RASIC) survey of members of biology and physics departments in Italian universities and research institutes measured response latency for each survey item. The RASIC dataset is a rich source of material. Respondent-level measures include extensive biographical data including age, academic rank, and language of choice (the survey was offered in Italian and English). 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, time of day, and device/browser used. The resulting dataset has respondent x item observations, with each observation being nested within respondent (e.g., age, tenure) and item (e.g., item length, reading grade level). Due to this nesting, a hierarchical cross-classified model is used for analysis. Our findings will shed light on the impact of a broad range of factors associated with response latency, addressing questions including the effects of time of day, age, means of access, reading grade level, number of response options, and so on. These analyses will provide important context for the perhaps simplistic interpretations of response latency: low latency being a desirable trait for items but undesirable for a respondent. Data collection utilized for this paper was funded by the Templeton World Charity Foundation, grant TWCF0033.AB14, Elaine Howard Ecklund, PI, Kirstin RW Matthews and Steven W. Lewis co-PIs.

Reducing Attrition Among a Sample of At Risk Teens

Tiffany Henderson, *Abt SRBI*
Dianne Rucinski, *Abt SRBI*
Kim Francis, *Abt Associates*
Michelle Woodford Martin, *Abt Associates*

Compared to cross-sectional surveys that are subject to nonresponse at a single point in time, longitudinal panel surveys suffer from both initial nonresponse at baseline and attrition - the loss of sample members in subsequent waves. Even when closely managed, attrition can reduce the statistical accuracy of estimates and result in attrition bias (Dajani, Bucholtz, & Warner, 2012; McCoy et al., 2009). Abt SRBI and Abt Associates completed a three year evaluation of a teen pregnancy prevention program among 1,644 teens in Minnesota. Students were randomly assigned to treatment and control groups across two cohorts and were surveyed at baseline, 12

AAPOR 70th Annual Conference

month and 24 month follow up periods. At the 12 month follow up, Cohort 1's attrition rate was unacceptably high (43%). Contact methods were adjusted to prevent and reduce attrition for Cohort 2, and limit subsequent attrition for Cohort 1. The survey window was moved up one month. Phone reminders were moved to the first week. The number of text and email contacts was reduced. Then, in addition to six and 18 month tracking, postcard reminders were added at four, 10, 16 and 22 months. As a result, the attrition rate for Cohort 2 prior to in-school administration was lower than Cohort 1 at both 12 (3% lower) and 24 month follow ups (5% lower). At the 24 month follow up, attrition only rose by 4% for Cohort 1 and only 1% for Cohort 2. In spite of our successful reduction of the attrition rate overall, our nonresponse analysis demonstrated that response propensity was associated with sample characteristics. Nonresponders to the 12 month survey across both cohorts were more likely to receive free or reduced lunch (64%), less likely to be Asian (14%), more likely to be Black/African American (31%), and more likely to engage in risky sexual behaviors.

Reducing Coverage Error in a Web Survey of College Students

Julie Pacer, *Abt SRBI*

Kelly Daley, *Abt SRBI*

While web surveys allow for fast and inexpensive data collection, they are subject to coverage error if respondents cannot be included because they cannot be identified or contacted (Groves et al. 2004). Coverage error is of particular importance in web surveys where often there is no listing of the population and their email addresses (Couper 2000). If a listing of email addresses exists, it may be incomplete or out of date. Firewalls or other security features may prevent an email from reaching even valid email addresses. In 2012-2013 Abt SRBI implemented the Survey of Scholarships in Science, Technology, Engineering, and Math (S-STEM) Recipients, a web survey to understand the effects of an undergraduate scholarship program on current and former students' college and work experience. The National Science Foundation (NSF) provided the sample which consisted of S-STEM recipients that received funding between 2006 and 2012. Most (97%) recipients had an email address listed in the sample file, but because the email address originated from the email given at the date of scholarship award, 55% of those addresses were four years old or older. In this research we describe efforts we undertook to verify email addresses in the sample and to improve contact rates. These efforts include telephone, mail, and internet based locating, commercial email location services, and alternating the sources of the email invitation between the study sponsor and the contractor. We report on the contact and completion rates by type of effort and total level of effort. We explore the relationship between coverage loss and certain sample characteristics such as student age, gender, major, and degree obtained as well as institution type (2 year or 4-year), region, and years since the scholarship was funded. The results will help optimize coverage in future web surveys of students with a listed sample.

Use of Continuation and Call-Back Interview Options to Increase the BRFSS Asthma Survey Response Rate

Barbara M. Fernandez, *Abt SRBI*

Marilyn Wilkinson, *Abt SRBI*

Dennis Daly, *Abt SRBI*

Andrew Burkey, *Abt SRBI*

Since 2005, the Behavioral Risk Factor Surveillance Survey (BRFSS) has fielded an Asthma Call-Back Survey (ACBS) for adults and children who currently have or formerly had asthma. Prior to 2014, the ACBS was conducted as a call-back interview. In 2014, Abt SRBI offered a

AAPOR 70th Annual Conference

second option to eligible respondents in Georgia, Maryland, New Jersey, and Ohio. In these four states, eligible respondents could choose the call-back ACBS interview or choose to continue immediately with the ACBS. While avoiding a call-back was expected to boost the response rate, the size of the boost would depend on the number of respondents willing to complete the longer interview in a single session. Preliminary results show that 62% of respondents completed the ACBS in 2014, compared to only 27% of respondents in 2013. Among those who completed the ACBS in 2014, 60% completed it immediately and 24% completed it as a call-back. We found no significant demographic or behavioral differences between those who completed the survey immediately after the BRFSS versus those we called back. An average of 1.01 call-back attempts were made in 2014 to complete an ACBS interview compared with 4.82 call-back attempts in 2013. This paper will report the results of the change in the ACBS methodology in terms of enhanced representativeness, response quality, and cost efficiency.

Comparing Results from Telephone Reinterview with Unmoderated, Online Cognitive Interviewing

William P. Mockovak, *Bureau of Labor Statistics*

This study compared results from a telephone reinterview with 30 employers involving 60 jobs with results from an unmoderated, online cognitive interview with ten participants to test the effectiveness of a revised set of questions designed to measure specific vocational preparation. Specific Vocational Preparation, or SVP, is defined as the amount of lapsed time required by a typical worker to learn the techniques, acquire the information, and develop the facility needed to achieve average performance in a specific job-worker situation. SVP is assessed using information about minimum education; required certifications, licensing, or specialized training; prior work experience; and job-related, post-employment training. Interviewer and observer debriefings from a previous small-scale field test had indicated that the questions used to measure SVP led to some respondent confusion and possible misreporting. Rather than wait for the next field test, a revised set of questions was developed and tested using reinterview and online cognitive testing. A reinterview with employers was conducted without reconciliation by two experienced interviewers (one conducting the interview, and one observing), and focused on identifying comprehension and communication problems. Business establishments were selected to provide a wide range of industries, geography, and size classes. The online cognitive interview was scripted and relied on participants talking aloud to determine how well the questions worked (only one question order was tested). Both pretesting approaches identified the same general problematic questions and issues, but the reinterview provided additional valuable information about a more effective reordering and revision of the SVP questions.

Contrasting Stylized Questions of Sleep with Diary Measures from the American Time Use Survey

Robin L. Kaplan, *Bureau of Labor Statistics*

Brandon Kopp, *Bureau of Labor Statistics*

Polly Phipps, *Bureau of Labor Statistics*

In the American Time Use Survey (ATUS), interviewers use a set of scripted, open-ended questions (Diary questions) to walk respondents chronologically through their activities during the prior 24-hour day, including sleep. The ATUS and other time diaries typically find that respondents sleep approximately 8.6 hours per night. Other surveys that ask respondents to report the average, normal, or typical amount they sleep per night (Stylized questions), find

AAPOR 70th Annual Conference

much lower sleep duration estimates, approximately 6.9 hours per night. We conducted 29 cognitive interviews to determine what might account for the discrepancies between Diary and Stylized sleep measures. Participants were asked both Diary questions about the previous day's activities and Stylized questions about sleep. Participants were then asked about their interpretation of the questions, their response strategies, and the social desirability of over-versus under-sleeping. As with the national studies, Diary questions led to longer sleep duration estimates than Stylized questions. Differences between the two estimates were tied to cognitive factors, such as participants' retrieval and estimation strategies. Participants with structured schedules either directly recalled how much they slept or used rate-based estimation techniques. Those with less structured schedules relied more on indirect cues (e.g., what television program they viewed before going to sleep). Differences between the two measures were also linked to social desirability concerns. Diary estimates closely matched participants' beliefs about what constitutes the maximum appropriate number of hours of sleep per night, whereas Stylized estimates aligned with what participants felt was a socially acceptable amount to sleep in one night. We identify possible sources of measurement error for Diary and Stylized sleep measures and implications for data quality and interviewer training. Follow-up research includes behavior coding transcripts of ATUS interviews to identify how respondents report sleep episodes, and contrasting self-reported versus objectively measured sleep via activity wristbands.

Only for the Young at Heart: Co-Viewing on Mobile Devices and Viewing on the Go?

Anh Thu Burks, *The Nielsen Company*

Jennifer Haskell, *The Nielsen Company*

Kumar Rao, *The Nielsen Company*

Helena Mendrisova, *The Nielsen Company*

With the relative ease and accessibility of a variety of content available to users of smartphones and tablets, there has been a subtle behavioral change in how people use these devices. The concept of viewing together or having more than one viewer for a mobile device is a phenomenon referred to as "co-viewing" and is a new area that warrants further investigation. Very little information is available on who is likely to engage in co-viewing behaviors, what types of mobile devices are used, what content is likely to be viewed and if those who engage in this activity / behavior are fundamentally different than those who are less likely—what are the behavioral or demographic differences among those who participate in these activities. Thus the focus here is to examine and provide a baseline understanding around the concept of co-viewing with specific focus of content viewing on the "go" or away from home. Lastly, a national representative survey was conducted where previous sampled respondents were asked a series of questions about if they watched video on mobile devices, how often they watch away from home, how often do others watch when away from home and the last time they shared with others when away from home. All respondents (N = 5,400) received a monetary incentive for completing the survey. Responses from these survey questions are discussed in terms of how these respondents differ (if any) from those who are less likely to engage in this behavior, the frequency in which co-viewing and out of home viewing occur, types of devices used and the frequency of sharing. Findings here and any difference noted, will provide a barometer of understanding of this type of behavior as it pertains to mobile device usage, location and media consumption.

English or Español? Examining the Relationship between Language Choice and Survey Quality among Bilingual Respondents?

Jennifer Haskell, *The Nielsen Company*

Anh Thu Burks, *The Nielsen Company*

Vera Kurmlavage, *The Nielsen Company*

Kumar Rao, *The Nielsen Company*

As the Hispanic population in the U.S. continues to grow, nationally representative surveys make an extra effort to ensure that the views of this important segment of the population are included. To better cater to Hispanics, surveys increasingly give respondents a choice of whether to conduct the survey in English or in Spanish in order to make sure that non-proficient English speakers are included in the sample. However, little is known about how survey language choice affects the quality of survey responses. To examine the effects of language choice on survey quality, we make use of a nationally representative RDD-based CATI survey (n=4,000) on mobile device usage that included a separate Hispanic sample. Respondents in the Hispanic sample, who made up almost one-tenth of the entire sample, were given a choice of whether to complete the survey in English or in Spanish. Bilingual individuals' choice of a language other than English in a survey represent an expression of their ethnic identity (Padilla and Perez 2003), which may cause them to respond differently than bilingual respondents who answer in English. In this study, we model the choice of survey language as a function of demographic and behavioral characteristics in order to understand the relationship between selected/modeled language choice and data quality. We compare and contrast the survey data quality among discordant respondents (those for whom modeled and reported choices were different) and concordant respondents in order to examine whether language choice drives differences between reported and expected behaviors. This article concludes with a call for further research on the role of language choice for bilingual respondents and recommendations for future research.

Doubling Down: Examining Survey Response Rates After Increasing Questionnaire Length

Amanda Libman, *The Nielsen Company*

Kelly Bristol, *The Nielsen Company*

Leah Christian, *The Nielsen Company*

As researchers continue to face declining survey response rates over the last decade, it has become increasingly paramount to understand how our questionnaire design choices may affect this decline. In July 2014, Nielsen needed to increase the number of questions in a mixed-mode survey for one of its TV Meter Panels, doubling the number of questions in the instrument from its first fielding in 2013 to its second fielding in 2014. These questionnaire changes were prompted by a methodological shift in part of the research's data collection. One major challenge from this change was to add the new questions need to meet the projects goals while avoiding a negative impact to our response rate. As survey researchers, it is often our job to balance operational needs and best practices of a project. A redesign of the questionnaire brought up discussions of rewording, page numbers, visual design, mixed-modes and operational restraints. When the new questionnaire was rolled out in July 2014, there was no significant drop in response rates, with a 2013 response rate of 20.6% and a 2014 response rate of 20.7%. This study will examine the changes made to the questionnaire, the lack of a change in the response rate, and the item non-response seen between the two questionnaires.

AAPOR 70th Annual Conference

The Paradox of Postcards: Examining the Effectiveness in Study Recruitment

Kay Ricci, *The Nielsen Company*

Lauren Walton, *The Nielsen Company*

Tracie Yancey, *The Nielsen Company*

Advance mailings before a survey are an industry best practice to establish legitimacy and connect with respondents (Dillman 2000). Many studies have demonstrated that sending an advance postcard is an effective means for increasing response rates (Hembroff et al. 2005). However, researchers have argued that a postcard can be overlooked, skimmed, or discarded within a matter of seconds, making both the material and its message forgettable. In an effort to better understand the effectiveness of postcards as a recruitment tool, Nielsen has conducted three separate studies to examine their impact on contact and cooperation. The findings of these studies were mixed, indicating that postcards can both help and hinder study participation. The first two studies found postcards to have a positive impact. Study 1 eliminated the advance postcard from mail recruitment, resulting in a significantly lower rate of survey returns (16.6% vs. 18.6%), and study 2 added a postcard in advance of telephone recruitment, resulting in directionally higher contact rates. Study 3, on the other hand, found an advance postcard to have a negative impact on study registration when compared to those who received nothing at all (9.6% vs. 10.4%). Further analyses will explore these findings in terms of the timing of mailings and the visual design of postcards, in an effort to better understand the contexts in which a postcard might prove beneficial. This paper seeks to inform the research community about the mixed effects of postcards in the recruitment process and their implications for best practices.

Crimes of Consolidation: Findings for Combining Household Communications

Lauren Walton, *The Nielsen Company*

Kay Ricci, *The Nielsen Company*

Jennifer Romano Bergstrom, *Fors Marsh Group*

Tracie Yancey, *The Nielsen Company*

Robin Gentry, *The Nielsen Company*

Survey organizations struggle to combat declining response rates without increasing costs. Each material sent to respondents should provide value enhancing their experience in a cost optimal way. In 2013, Nielsen partnered with Fors Marsh Group to qualitatively assess the impact of materials sent to homes during Nielsen TV diary studies, which uses a two-stage mixed-mode design on an address-based sample. The ethnographic study was conducted over video chat after mailing materials to participants. At the diary stage, a letter, question and answer brochure (Q&A), cash incentive, and diary are sent to households. The ethnographic study suggested a lack of unifying path in the respondent experience (i.e., not all participants looked at the materials in the same order). As a reaction to this, Nielsen tested a "Quick Start Guide" approach, combining the letter and Q&A in July 2014. The goal of this material was to unify the experience across respondents and to provide needed information in a simple way. Analyses indicate a significant negative impact to cooperation (returning a usable diary) using this Quick Start Guide approach (28.0% vs. 27.4%). Data quality suffered in the test group. The Quick Start Guide test group returned significantly more unusable/bad diaries (17.9% vs. 19.9%). These results suggest consolidating materials may save money (in this case, one cent per mailing), but can negatively affect key metrics.

AAPOR 70th Annual Conference

Text That: SMS for Survey Data Collection in Developing Markets

Jeff Scagnelli, *The Nielsen Company*

Jacques Human, *The Nielsen Company*

Janice Linnane, *The Nielsen Company*

The growth in mobile access across developing markets has enabled a new wave of data collection methodologies for survey research. The ability to access new channels of communication and services is the driver for the growing adoption rates. The growing ownership is not limited to higher socioeconomic groups, which actually now allows access to traditionally hard to reach populations who are less accessible through traditional means. In countries like Kenya, the household level presence of copper-wired landline telephones is a fraction of the mobile phone penetration. The drivers for this beyond cost are high maintenance costs, as well as theft of the copper wire itself. These challenges, along with the desire to better understand these marketplaces place an emphasis onto the survey research community to utilize mobile based data collection methods, like SMS. This paper contributes to the literature around SMS data collection for consumer product survey research. This pilot took place in February 2014 in Kenya. Users were asked to report their beverage purchases through a daily survey. Households would respond to the SMS based prompts with the details of their purchase. In exchange for their reporting, they were provided with mobile airtime credits. The main objectives for this work were to assess the viability of this approach, as well as analyze the ability to re-engage the same users a second time. These results will add to the knowledge base for conducting survey research in emerging markets.

Estimating Store Sales Using Photography

David Quach, *The Nielsen Company*

Lukasz Chmura, *The Nielsen Company*

Choongkoo Lee, *The Nielsen Company*

Victoria Tsay, *The Nielsen Company*

Photography is currently used by Nielsen and other market intelligence companies to gather information about retail stores such as where products are placed in a store and whether a product is out of stock. A natural extension to the use of in-store photography for these purposes is to use photography to estimate sales. Nielsen tested the use of photography to estimate the sales of 10 products in grocery stores in the Chicagoland area. We found that actual sales were not significantly different from estimated sales for four of the ten products tested. The sales of the other six products were estimated with a relative mean error of 6 to 24%.

Validation of Metrics: A Comparative Analysis of Predictive- and Criterion-Based Validation Tests in a Qualitative Study

Erin Michele Fordyce, *NORC at the University of Chicago*

Sabrina Bauroth, *NORC at the University of Chicago*

Catherine Vladutiu, *Maternal and Child Health Bureau, Health Resources and Services Administration*

A primary concern for researchers, when collecting self-reported data from respondents, is supporting the accuracy and reliability of the data collected. Measurement error can be introduced by questions asking for specific, factual information that may be difficult for respondents to recall. It can further increase due to the underreporting of sensitive information

AAPOR 70th Annual Conference

(Blattman et al., 2014) often attributed to social desirability. Researchers continue to struggle with finding cost-efficient and burden-free approaches to validate self-reported information. The National Survey of Children's Health (NSCH) collects data on the physical and emotional health of children including current and pre-existing conditions (or disabilities). As part of the NSCH Redesign study, 90 cognitive interviews were conducted with predictive and criterion-based validation tests to validate household screener items as well as items related to medical diagnosis and health insurance status. To assess predictive validity, a test-retest approach was used, whereby a subset of respondents were re-administered items from the household screener and main questionnaire two weeks subsequent to their initial interview. A separate, criterion-based validity test was conducted by asking the remaining respondents to provide documentation to validate the household screener items as well as items related to medical diagnosis and health insurance status. Respondents who did not provide documentation were asked permission to contact the child's primary care provider. This paper addresses important issues surrounding instances where a researcher is asking respondents to provide documentation to validate reported information. For instance, we discuss the impact of requesting various types of documentation on respondent burden and the advantages and disadvantages of requesting documentation versus other means of validating information. In addition, we discuss the effectiveness of conducting retests in identifying potential measurement error.

Leveraging Area Probability Sampling in Recruiting Households for Web Surveys

Steven Pedlow, *NORC/University of Chicago*

Al Tupek, *NORC/University of Chicago*

Kennon Copeland, *NORC/University of Chicago*

Probability-based recruitments for web surveys typically use address-based sampling or (in the past) random digit dialing methodology for sourcing their samples. An alternative approach is to source the sample primarily from area probability sampling and then supplement the sample, as needed, from addressed-based sampling (ABS). This alternative approach has the benefit of improving sample coverage. In addition, the use of area probability-sampling with ABS supplements facilities the cost-effective use of in-person interviewing to improve the response rate and minimize bias. The paper will present a case study of the application of this alternative approach for supporting the sample requirements for recruiting households to participate in web surveys. While probability-based sampling for web surveys is desirable on many levels (e.g., sample representativeness, ability to calculate standard errors, less bias in the estimates), low response rates and sample noncoverage remain a challenge to recruiting representative samples for web-based surveys. The application of this alternative approach is useful for any potential study where it is desirable to obtain the highest possible response rate and optimal sample coverage for recruitments of surveys. The paper will describe the considerations in using an area probability approach based on a case study using the NORC area probability sample frame, including a discussion of sampling in rural areas where United States Postal Services addresses can be supplemented by in-person listing. The paper describes the techniques used to replace some area probability clusters using ABS.

AAPOR 70th Annual Conference

RTI Mobile Maps Application for Field Surveys

Katherine Morton, *RTI International*

Charles Loftis, *RTI International*

Bonnie Shook-Sa, *RTI International*

In area household surveys, field staff use maps to locate sample dwelling units and to apply a frame supplementation procedure such as the half-open interval or check for housing units missed. Maps are also used to build field enumerated (FE) frames by recording the address and location of each dwelling unit within the sampled area. Replacing traditional paper maps with electronic maps leads to time- and cost-savings and increases capabilities for managing field work. RTI International (RTI) recently developed a cost-efficient, state-of-the-art mobile mapping application that can support surveys of many sizes and frame types (FE, address-based sampling, etc.). The user-friendly application was developed using open-source software and can run on multiple operating systems and devices. We present the features of the RTI Mobile Maps application including the ability to add dwelling units and roads to existing maps, transfer data to and from a mobile device, and re-integrate new content back into a database. We also present the results of a small pilot test in which field staff enumerated dwelling units using the RTI Mobile Maps application and provided feedback.

Fit for Purpose: An Enhanced Quality Perspective for the National Internet Flu Survey

Michael Bostwick, *RTI International*

Jill A. Dever, *RTI International*

M. Christopher Stringer, *RTI International*

Tammy A. Santibanez, *Centers for Disease Control*

Anup Srivastav, *Centers for Disease Control*

Stacie Greby, *Centers for Disease Control*

Peng-Jun Lu, *Centers for Disease Control*

Mansour Fahimi, *GfK Custom Research LLC*

Michael Lawrence, *GfK Custom Research LLC*

Given the limitations of traditional survey designs, assessment of alternative designs calls for an understanding of data quality as a multidimensional concept that addresses the fitness of the data for its intended purpose (Biemer & Lyberg, 2003). Specifying a survey's 'fit for purpose' is an important initial step towards ensuring the high-quality results meet the intended needs through an agreed-upon definition. Six example 'fit for purpose' criteria include relevance, accuracy, timeliness, accessibility, interpretability, and coherence. For example, rapid implementation of surveys and dissemination of high-quality results are critical for surveys designed to capture policy-relevant, time-sensitive population estimates. The Centers for Disease Control (CDC) in collaboration with RTI International and GfK Custom Research LLC (GfK) used the 'fit for purpose' paradigm to develop and implement such a critical-needs survey. CDC designed the National Internet Flu Survey (NIFS) to estimate early-season vaccination coverage rates and knowledge, attitudes, behaviors, and barriers related to influenza and influenza vaccination in the U.S. adult population [relevance]. Data for the 2014-15 NIFS was collected over a two-week period [timeliness] via computer-assisted web interviews (CAWI) from a stratified random sample of noninstitutionalized adult members of the U.S.-based GfK Knowledge Panel [accessibility]. Allocation of the sample to the design strata was determined based on a set of precision goals for the population estimates [accuracy]. The CDC used the 2014-15 NIFS data to generate population estimates for adults 18 years and older on early-

AAPOR 70th Annual Conference

season vaccination estimates for dissemination [interpretability] during the National Influenza Vaccination Week, December 7-12, 2014.[1] We will detail the 'fit for purpose' attributes of the 2014-15 NIFS and where possible compare the characteristics to other surveys including previous rounds of the NIFS [coherence]. [1] <http://www.cdc.gov/flu/nivw/>

Effect of the Mode of Collection in Statistics Canada's Fuel Consumption Survey

Agnes Waye, *Statistics Canada*

Serge Godbout, *Statistics Canada*

Pierre Daoust, *Statistics Canada*

The National Fuel Consumption Survey (FCS) was created in 2013 and is a quarterly survey that is designed to analyze distance driven and fuel consumption for passenger cars and other vehicles weighing less than 4,500 kilograms. The sampling frame consists of vehicles extracted from the vehicle registration files, which are maintained by provincial ministries. For its mode of collection, FCS uses car chips for a part of its sampled units to collect information about the trips and the fuel consumed. There are numerous advantages to using this new technology, for example, reduction in response burden, collection costs and effects in data quality. For the quarters in 2013, the sampled units were surveyed 95% via paper questionnaires and 5% with car chips, and in Q1 2014, 40% of sampled units were surveyed with car chips. This study outlines the methodology of the survey process, examines the advantages and challenges in processing and imputation for the two collection modes, measures the mode effect and concludes with a summary of the lessons learned.

Challenges in Developing a New Collection Strategy for the Canadian Community Health Survey (CCHS) Redesign

Marie-Claude Duval, *Statistics Canada*

The Canadian Community Health Survey (CCHS) is a cross-sectional survey that collects general health information on the Canadian population at the community level. This survey is undergoing a redesign that will take place in 2015. Among the changes, a new collection strategy has been developed using a multi-mode approach. While one of the goal was to reduce the collection cost, the survey had to delay its expectation in order to maintain quality with the use of the new Household Survey Frames Service (HSFS) available at Statistics Canada. In the long run, a reduction in collection costs is expected with the continuing improvements to the HSFS and the introduction of an Electronic Questionnaire (EQ) when the new Generalized Collection System at Statistic Canada is available. The presentation will describe the issues that were faced while determining a new collection strategy, the multi-mode collection strategy adopted in the CATI and CAPI environments for 2015 and the expected changes for the coming years. The presentation will also briefly cover the use of responsive design to control and balance data quality and collection cost.

Experiences in Improving Response Rates for Household Surveys

Sylvie Bonhomme, *Statistics Canada*

As many statistical organizations, Statistics Canada has observed a downward trend in response rates for household surveys due to changes in the external and internal environment. To deal with this issue, Statistics Canada has developed a research framework that looks at numerous viable options available through out the survey collection cycle and tries to identify strategic points in the course of the collection process where initiatives could be implemented to

AAPOR 70th Annual Conference

address this issue. Initiatives such as using better communication strategies with potential respondents, introducing the use of incentives, making more household surveys mandatory, improving frame coverage and improving data collection process and practices are being considered, analyzed and/or tested. Other important initiatives being brought forward are the introduction of a collection dashboard using key indicators to be used in the field during collection in order to provide a better decision-making tool to manager in the course of collection. The introduction of responsive collection design in collection for all household surveys and a better use of survey paradata are also being initiated. Findings resulting from recent surveys have identified many options and opportunities for improvements. In particular highlights of a recent non-response follow up survey generated many ideas of research projects and potential experiments that can be used to both improve data collection process and response rate. Results from the introduction in surveys of multi-mode collection focusing especially on electronic questionnaires is also being scrutinized as well as the increase use of cellular phone by respondents and therefore in the collection process. The main objective of this poster session is to presents and share results of initiatives being implemented, present some of the findings of those initiatives and promote discussions and share experiences about potential and operationally viable initiatives, process and measures that could be implemented to improve survey response rates.

The Redesigned Canadian General Social Survey: Our First Experience in a Multi-Mode Collection Environment

Pierre Caron, *Statistics Canada*

Marie-Hélène Miville, *Statistics Canada*

Patrick St-Cyr, *Statistics Canada*

The General Social Survey (GSS), established in 1985, conducts annual cross-sectional surveys across the ten Canadian provinces. The GSS is recognized for its regular collection of data that allows for trend analysis, and its capacity to test and develop new concepts that address current and emerging issues. Historically, the GSS has been a CATI survey that has relied on random digit dialling (RDD) methodology for its sample design. Increasing coverage issues with RDD and the introduction of new collection methods such as electronic questionnaires for internet collection were key factors leading to a redesign of the GSS in 2013. The presentation will first focus on the sampling frame options considered as replacement to RDD methodology. Both address-based and telephone-based approaches were evaluated and will be discussed. In addition to improved coverage, a desirable property of the new frame is to have auxiliary information that may be used at the design stage, such as information to target specific subpopulations. The 2013 GSS on Social Identity, the first under the redesign, also marked the first time a social survey at Statistics Canada offered an Internet option to survey respondents. This new approach to data collection was in recognition of the success of online data collection with the Census, combined with the need to adapt to the changing use of technology and the ever present demands on Canadians' time. We'll describe the GSS multi-mode methodology offered to respondents for GSS and discuss major advantages and disadvantages of the implemented method. Finally, we'll present collection results from our first cycle under the redesign. More precisely, we will present and explain the various factors that have lead to a reduced response rate under the new design.

Assessing the Validity of Two Methods of Collecting the Number of Rooms in Housing Units: Is there a Measure of Truth?

Mikelyn V. Meyers, *U.S. Census Bureau*

Dawn Nelson, *U.S. Census Bureau*

In an effort to improve data quality, in 2014 the U.S. Census Bureau conducted a split ballot field test to determine which of two methods more accurately collects the number and types of rooms in housing units for the American Housing Survey (AHS). Collecting a count of rooms has been hampered by the discrepancy between the Department of Housing and Urban Development's administrative definitions for rooms and respondents' definitions of rooms, which are based on room usage, the standards set by realtors, building codes, and health codes. The field test compared the performance of a room-based approach of collecting room counts to a floor-based approach believed to more closely conform to how respondents use visual imagery to retrieve information. Data was available for each housing unit from multiple sources, including longitudinal data from a prior administration of the AHS, respondents' answers to the field test survey questions, observational data from a tour of the home conducted by interviewers, respondents' answers to debriefing questions, the results of behavior coding the survey questions, and public tax records. While the intent of the study was to compare respondents' answers to the survey questions with both the inventory of rooms from the tour of the home and with public records in order to determine which method generally yielded a more accurate room count, establishing a measure of "truth" proved to be problematic, as both the tour of the home and the public records had noteworthy limitations. We will address how both candidates for a measure of truth fared when compared to the different types of data that were collected, and how, in light of the subjective nature of determining "truth" for this field test, one method of collecting the number of rooms in housing units was ultimately determined to be more accurate than another.

Interviewer Compliance and Data Accuracy: Evidence from the Field

Mandi Martinez, *U.S. Census Bureau*

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

In an effort to improve data quality, the U.S. Census Bureau conducted a split ballot field test in 2014 to determine which of two methods more accurately collects the number and types of rooms in housing units for the American Housing Survey (AHS). The first method followed the current room-by-room inventory approach and the second explored a new floor-by-floor approach believed to more closely conform to how respondents use visual imagery to retrieve information. A team of experienced and newly hired interviewers administered both versions of the room inventory questions using a paper questionnaire. With respondent consent, the interviews were tape recorded. At the end of the interview session, interviewers conducted a home tour to observe and record the number and type of rooms in the housing unit. To assess room count accuracy, the room counts from both questionnaire versions were compared to the room counts collected in (1) the home tour and (2) supplementary public records data. When testing question wording, an underlying assumption is that interviewers read the questions as worded. While monitoring in real time, it became apparent that our interviewers were not complying with this assumed behavior. Further, the interviewer compliance rate varied across (1) interviewers, (2) questionnaire methods and (3) housing types. In this poster we report on the extent of interviewer non-compliance across the three dimensions; we introduce potential reasons for the non-compliance; and, finally, we discuss the associated effect of non-compliance on room count accuracy.

AAPOR 70th Annual Conference

Distribution of Characteristics by Mode in a Sequential Multi-Modal Survey

David A. Raglin, *U.S. Census Bureau*

Sequential multi-mode surveys are known to attract different types of respondents in each mode--but what differences do we see in the characteristics of those who choose to respond in each mode? Although it is difficult to identify the causes, whether they be mode effects or respondent preferences, it is important to understand the differences in the people who respond by each mode. The American Community Survey (ACS) is a sequential multi-modal survey conducted by the Census Bureau which produces data previously available from the Decennial Census long form. The ACS collects data in a series of monthly panels using four modes of data collection over a three month period. In the first month, we mail sampled addresses an invitation to respond via the Internet, with nonresponding addresses later mailed a paper questionnaire. In the second month, we call addresses for which we have a telephone number. In the third month, a sample of the remaining addresses who have not responded are followed up using personal visit. This paper examines the demographic, social, and economic characteristics of the households that respond using each of the four data collection modes. Some of the characteristics are race, Hispanic origin, household size, owner/renter, language spoken at home, education level, health insurance coverage, employment status, and poverty. The paper shows how the characteristics of the households that respond to each mode differ, and it discusses reasons for the differences, as well as the implications for the ACS and other surveys.

What are the Effects of Proposed Changes to the American Community Survey Internet Instrument?

Mary Frances E. Zelenak, *U.S. Census Bureau*
Rachel T. Horwitz, *U.S. Census Bureau*

Starting with the January 2013 data-collection cycle, the U.S. Census Bureau began collecting American Community Survey (ACS) response data via an Internet instrument. Before going live in production, the Internet instrument went through several rounds of usability testing and was used in the April and November 2011 ACS Internet Tests. In those tests, we found that higher item nonresponse rates for items later in the survey were related to Internet breakoffs. From the paradata, we also saw that screens containing questions with multiple parts (i.e. a check box or radio button response followed by a write-in response) tended to render error messages frequently due to respondents missing the associated write-in box. Based on these findings, as well as feedback from respondents who were unable to return to the Internet survey because they had forgotten their assigned User ID and PIN combination, the Census Bureau proposed changes to the ACS Internet instrument to make it more user-friendly and reduce breakoffs and item nonresponse rates. This poster will discuss the specific changes and the effects of these changes as seen in the 2014 ACS Internet Breakoffs Test.

Are You Still There? Using Respondent-Provided Email Addresses to Send Invitations for an On-Line Survey

Jonathan Hoechst, *Tetra Tech*
Mandy Pom, *Tetra Tech*
Peg Kreckler, *Tetra Tech*

Survey researchers increasingly use, or would like to use, email addresses to invite sample members to complete on-line surveys. While postal invitations to complete on-line surveys can

AAPOR 70th Annual Conference

yield acceptable response rates and overcome significant coverage issues associated with email lists, issuing survey invitations by email further reduces costs and makes reminder notices more efficient. However, relatively little is known about the quality of the email lists and respondents' reactions to emailed invitations. Researchers may learn very quickly that an email address is "bad" or receive confirmation the address is "good" when the survey is completed, but this leaves a vast gray area of potentially valid addresses that are not actively used. For a large, annual, nationwide survey that will be conducted in early 2015, we will use email addresses that were provided by respondents over the previous four survey years to send invitations to complete the on-line survey (approximately 25,000 email addresses). We will track and analyze the rate at which survey invitations bounce-back (confirmed "bad"), result in a completed survey (confirmed "good"), as well as yield confirmation of potentially valid addresses (email not opened, email opened but survey link not clicked, or survey started but not completed). To the extent possible, we will also conduct these analyses by age of the email address and the type of domain or provider. The analysis, while a "best-case scenario" based on respondent-provided email addresses, will provide useful information on the quality of email frames and the stability of this information over time.

A Randomized Pilot Study of Three Approaches to Increase Participation in the GuLF STUDY Follow-Up Interview

Polly P. Armsby, *Social & Scientific Systems, Inc.*

Matthew D. Curry, *Social & Scientific Systems, Inc.*

Carley L. Prynn, *Social & Scientific Systems, Inc.*

Ryan J. Chaffee, *Social & Scientific Systems, Inc.*

John A. McGrath, *Social & Scientific Systems, Inc.*

Richard R. Kwok, *National Institute of Environmental Health Sciences*

Larry S. Engel, *School of Public Health, UNC-Chapel Hill*

Dale P. Sandler, *National Institute of Environmental Health Sciences*

The Gulf Long-Term Follow-up Study (GuLF STUDY) is a prospective study of health effects among 31,362 adults who worked or sought work in the clean-up effort following the 2010 Deepwater Horizon oil spill in the Gulf of Mexico. Participants were enrolled between March 2011 and March 2013 and follow-up interviews began in March 2013. Some characteristics of the cohort make them more challenging to follow. For example, 29% report household income below \$20,000 and 16% did not graduate from high school. A high response rate is needed to avoid bias. Our initial approach to promote participation included advance letters, reminder mailings, batch tracing to obtain contact information updates, gift card drawings for responders, and refusal conversion efforts. Yet, participation was lower than desired. As of July 2014, 59% of participants had been contacted, and 86% of those reached completed an interview. Uncontacted participants had telephone numbers that were disconnected (51%), rang unanswered (29%) or were no longer correct (18%). We pilot tested three methods to increase follow-up. 150 participants were randomized to one of three interventions: mailed \$25 gift card, mailed pre-paid cell phone, or pre-paid cell phone delivered by field locators. Following these interventions, we placed additional calls to working telephone numbers. Direct locating and delivery of cell phones resulted in more completed interviews (10%) than mailed gift cards (8%) and mailed cell phones (6%). Because mailed gift cards were nearly as effective as direct delivery of phones and considerably less expensive, we sent mailings in September and October 2014 to all remaining non-responders offering a \$25 gift card for completing the interview. As of early November 2014, 5% had contacted the study center to complete a telephone interview. To further maximize participation, additional field-based locating is planned for targeted subgroups.

Will Personalized Graphics Help Improve Response Rates in Distrustful Populations?

Morgan S. Jones, *East Tennessee State University*

Stephanie P. Elliott, *East Tennessee State University*

In Fall of 2014, all residents and business owners in the community of Roan Mountain, TN have been asked to participate in a survey about the possible installation of a public sanitary sewer system since their community currently does not have one in place. Roan Mountain is a very rural area in Northeast Tennessee that is, at best, ambivalent about research but often outright suspicious of research efforts. This is a significant concern when thinking about response to a survey and non-response bias because the responses to this survey will serve to formulate policy and allocate funds in the community. Because this community is so small, it is critical to gain the input of as many residents and businesses as possible. Therefore, we have designed an experiment to see which invitation strategy will have more of an impact; a generic, university branded envelope or one that is branded ETSU but also calls on community affiliation by having a picture of the Roan Mountain community and a message designed to try and evoke hometown pride. The respondents for this study have been randomly divided into two groups and response rates and possible non-response bias will be tracked for each experimental condition. The population for this study is comprised of 160 businesses and 2,294 residents, which is the entire population of businesses and residents of this community. The survey is a mail survey and all contacts will be branded with their experimental condition. It is our hope that this project will shed light on ways to increase the response rate in communities with historically low participation rates.

Money Talks: Purposeful Incentive Increases in a Longitudinal RCT

Ronald E. McCowan, *Decision Information Resources, Inc.*

Sylvia R. Epps, *Decision Information Resources, Inc.*

Ronald H. Bass, *Approximetrix*

Achieving sufficient response rates among hard-to-reach samples present challenges, especially among longitudinal experimental evaluations. We know, through empirical evidence, that offering incentives can be an effective strategy towards ensuring and improving response rates (Singer, Groves, and Corning, 1999; Groves, Singer, Corning, and Bowers, 1999). Still, there is no 'one size fits all' solution and questions remain about the cost-effectiveness of incentive increases, their optimal timing, or their influence over-time with longitudinal samples. The proposed poster will focus on two waves of data from an experimental evaluation, which included planned incentive increases in each wave. Specifically, data from the SIF SaveUSA evaluation will be used. This evaluation included low- and moderate-income tax filers who received services at Volunteer Income Tax Assistance (VITA) sites in four states during 2011. MDRC and DIR designed and implemented an RCT evaluation of SaveUSA in New York City and Tulsa, and completed data collection 18 months and 36/42 months post random assignment, achieving 80% response rates each wave. The second wave included both first wave survey completers and non-completers. At 18-months, the incentive started at \$25 and increased to \$50; at 36/42-months, the incentive started at \$30 with increases to \$60 and then to \$100. Using data on each respondent's incentive amount for each wave, we will address the following: ·Does a higher incentive at Wave 1 impact responsiveness at Wave 2? ·Does a higher incentive during Wave 2 differently impact responsiveness of first-wave completers and non-completers? Data from approximately 1400 households will be analyzed to address these research questions. Preliminary analyses indicate that respondents receiving the lower incentive

AAPOR 70th Annual Conference

amount at Wave 1 were not less likely to complete at Wave 2. Ultimately, the findings will improve knowledge around data collection planning to assist researchers setting or increasing monetary incentives, particularly on multi-wave studies.

Your Money's No Good Here: Who Returns a Prepaid Incentive?

Ashley Kaiser, *AIR*

Danielle Battle, *AIR*

It's well known that incentives increase response rates in surveys. Research suggests that prepaid cash incentives are most effective (Church, 1991; Singer & Kulka, 2004). They can have a positive effect on the effort that respondents are willing to put into a survey, increase participation of underrepresented groups, and decrease overall survey costs (Singer, et al, 2004). Projects budget for an incentive for each sample member, regardless of their participation - but what happens when the incentive is returned? In 1960, Project Talent (PT) collected extensive cognitive, personality and background information from 377,000 9th-12th graders. In 2014, fifty-four years later, the American Institutes for Research (AIR) launched The Project Talent Twins and Siblings Study (PTTS). PTTS is a sample of all twins and their siblings who participated in Project Talent in 1960 (n=3,924). PTTS sample members were mailed a survey package that included a ten dollar bill. Although most people completed the survey (n=2,325), those who refused (n=315) did so in various ways (e.g., they called the telephone hotline, sent an email, sent back a blank survey, or wrote a note). Of the 315 refusals, 122 people also returned the \$10 incentive in the pre-addressed, pre-stamped return envelope provided in the survey package. Additionally, 18 sample members who completed their survey also returned their incentive. Using PT, this paper will examine the demographic and personality characteristics of those who return their incentive. We will look at how those who return incentives differ from other response and refusal groups. Results could inform responsive design for future collections (e.g., Could we incentivize groups with similar characteristics differently to increase response rates? For example, if a subset scores very high on measures of altruism, should we consider donating to a charity for that subgroup rather than mailing the incentive directly to them?).

Differential Reporting of Administrative Record Data by Cell Respondent Location

Becky Reimer, *The Associated Press-NORC Center for Public Affairs Research*

Dan Malato, *The Associated Press-NORC Center for Public Affairs Research*

Christopher Ward, *NORC at the University of Chicago*

Jennifer Benz, *The Associated Press-NORC Center for Public Affairs Research*

Jenny Kelly, *NORC at the University of Chicago*

Trevor Thompson, *The Associated Press-NORC Center for Public Affairs Research*

Telephone survey research methods have adapted in recent years as U.S. households have shifted away from landline telephones and increasing proportion of households have become reachable only or mainly by cell telephones. The inclusion of cell telephone frames is now nearly essential for coverage reasons, but cell interviewing presents researchers with distinct challenges. Recent research (Reimer, et al., 2014; Lavrakas, et al., 2010) suggests that more than a quarter of cell phone respondents who complete interviews do so while in locations away from home. These respondents may be less willing and/or able to provide certain types of information than landline respondents or cell respondents who are at home when interviewed. This difference may be magnified when respondents are asked for administrative record information, including when respondents are asked for consent to obtain medical records (Ward, et al., 2013) and asked to give contact information for medical providers to obtain those records

AAPOR 70th Annual Conference

(Ward, et al., 2014). The present study extends previous research using a survey where respondents were asked for account and billing information so specific that most would be unable to recall without consulting physical documents or other records. Analyzing data from this survey about household utilities, we examine the behavior of cell respondents who were at home vs. away from home when asked to provide specific details from their bills like amounts and usage, as well as consent to contact service providers using account numbers to obtain a fuller picture of their behavior. We assess differences in cooperation rates between these groups, and discuss cost and effort implications of callback strategies that could be employed to equalize differences in obtaining personal record information.

Going It Alone: Experiences of a Volunteer Helping a Not-for-Profit Organization to Conduct a Survey

Karen L. Goldenberg, *Retired*

Retirees are often encouraged to find places to volunteer. This poster reports on my experience as a volunteer, drawing on my professional background in survey research to assist a small not-for-profit organization to develop and conduct a mixed-mode membership survey. The task was in significant contrast to my prior work in a Federal statistical agency. There, I had access to extensive technical and research support; here I was the only person on the project familiar with the survey process. Challenge 1: Piecemeal development. The original focus was on the questionnaire, not the survey as a whole. The intent was to collect data electronically where possible, but within the organization there was little understanding of the process for doing so. Over time, and with my input, discussions evolved into the need for online survey software, recommendations for advance notification to email respondents, and a mail questionnaire for the members for whom we had no email addresses.. Challenge 2: Going it alone. I had the knowledge and expertise to design and test a paper questionnaire. But in the absence of a support organization, it was also up to me to lay out a timeline, learn how to use the survey software, coordinate online and paper versions of the instrument, and understand how the email/data capture process within the software works. The initial mailing is planned for mid-November 2014. The poster will address these and other challenges, including: - Differences in priorities, e.g., getting “enough” responses versus a high response rate - Limited staff resources - Timing of data collection - Source of decision-making authority - Concerns for both the organization and volunteer from reliance on a single non-staff person. Some substantive results will be reported.

Examining Survey Response with a “Choose Your Own Adventure” Approach

Tami Buhr, *Opinion Dynamics*

Melanie Munroe, *Opinion Dynamics*

The survey research industry is at a crossroads. Telephone survey response rates have been declining for the past quarter century. The increased use of cell phones and call screening devices have accelerated this trend. Many people simply will not answer their phone if the call is from an unknown number. However, these same people may respond to a request to complete a survey that they receive through the mail or email. In this environment, multi-mode fielding practices that allow the respondent to complete the survey through their preferred mode may be beneficial to survey response. We will present results from a survey experiment conducted in 2014 that utilizes different survey fielding modes and incentive types. The objective of the experiment is to compare the response rates and results of a traditional telephone survey with one that makes use of multiple survey modes and incentive types. For this experiment we use a split-sample design in which we conducted a traditional telephone survey with half of the

AAPOR 70th Annual Conference

respondents and a multi-mode survey with the other half. For the multi-mode survey, we mailed the survey invitation and respondents were given the option to complete the survey online through an enclosed web link or a telephone number to call to complete the survey with an interviewer. In addition, we varied the survey incentive structures used to encourage participation providing an up-front small incentive for half the multi-mode respondents and a sweepstakes entry for a much larger prize for the other half. The traditional telephone survey respondents were all provided a sweepstakes entry. Across all experimental conditions, we will compare the response rates, demographics, and survey responses. We also explore the cost implications of the different survey modes and incentive structures.

Mode Effect for Minimally Invasive Questions Compared to Invasive Questions

Sari E. Schy, *NORC at the University of Chicago*

Alyssa Ghirardelli, *NORC at the University of Chicago*

This poster presents analysis of questions that are minimally invasive questions (income, age and race) and one question that is an invasive question (report of having a mental illness now or in the past). It compares responses to four questions across three different modes of data collection and identifies if mode effect exists. Previous research shows that, when discussing mental illness, mode effect does exist. This poster will determine if that is also true for this study. The three different modes of data collection were: 1. Interviewer administered telephone screener, 2. Web self-administered questionnaire, and 3. Self-administered pen and paper questionnaire. The subjects were initially contacted via telephone by recruiters and given a screener that the recruiters filled out. After the subjects completed the screener, were identified as eligible and consented to take part in the study, they were emailed a link to a web-based survey that was confidential. The subjects filled out the online questionnaire approximately one week prior to attending a large in-person focus group. The online survey was used to reduce response bias such as a demand effect that could occur from a pre-test administered just prior to exposure and a post-test immediately following exposure. At the in-person session, subjects watched a video about reducing mental health stigma and discrimination and engaged in group discussion. After the discussion, the subjects completed a self-administered pen and paper questionnaire. Data from all three modes of data collection were merged and analyzed.

How Much Do You Hate the Other Guy? More So On the Web Than On the Phone

Jeffrey Gottfried, *Pew Research Center*

Ruth Igielnik, *Pew Research Center*

For years, favorability ratings have often been used to measure the public's assessment of political figures. The rapid growth in the use of self-administered web surveys has led to exciting new avenues of measurement as well as challenges in how we interpret the public's opinion toward our political leaders. As self-administered surveys become more common, one of these challenges is to understand the influence of survey mode on how respondents rate political figures. This presentation will provide an in-depth analysis of size and direction of mode differences about highly visible political figures, including Hillary Clinton, Michelle Obama, George W. Bush, Mitch McConnell, Harry Reid, and Sarah Palin. Additionally, since the magnitude of mode differences can vary by subgroups, particularly when a topic is sensitive or cognitively difficult, we analyze how mode effects vary by respondent party identification. Respondents for this study are drawn from the Pew Research Center's American Trends Panel, recruited in early 2014 in an RDD telephone survey. The experimental groups include 1,509 respondents who took the survey on the web, and 1,494 who completed a telephone survey. Findings show that, overall, respondents who identify with the opposite party as the political

AAPOR 70th Annual Conference

figure are substantially more negative in their evaluation over the web than on the phone. Alternatively, respondents who identify with the same party as the political figure do not differ in their favorability between modes. Thus, differences in modes on the assessment of political figures are mediated by in-group and out-group party affiliation.

Recruitment into a Clinical Trial: Comparison of Two Online Sample Sources

Jordon Peugh, *SSRS*

Ellen Meier, *Medical University of South Carolina*

Amy Boatright, *Medical University of South Carolina*

Amy Wahlquist, *Medical University of South Carolina*

Matthew Carpenter, *Medical University of South Carolina*

Survey research panels can be used for clinical trial recruitment, but not without specific challenges. Lengthy exclusion criteria and non-response at each consenting step often require large numbers of potential participants to screen. For a recent clinical trial of smoking cessation, focused exclusively on smokers unmotivated to quit, we used two online panel sources to recruit participants. We first selected a probability-based sample in an effort to increase confidence in the representativeness of our data. However, since this sample was not large enough to meet our needs, we secondarily used online convenience sample. We herein focus on potential differences between study participants that were recruited through these two sources. Between 11/2011 and 8/2013, we screened, recruited, and consented 1,233 study participants: 314 from the probability-based panel and 919 from the convenience panel. Recruitment samples did not differ on most demographic variables (including age, gender, race, Hispanic ethnicity, marital status, and income). The probability-based sample was less educated and more likely employed full-time compared to the convenience sample. Looking at critical study variables, there were no significant differences in number of cigarettes smoked, age at smoking initiation, number of quit attempts, nor current smoking status. There were also no differences in study retention (number of scheduled contacts completed) between recruitment sources. In our full paper, we will also examine differences in eligibility and consent rates between the two samples and discuss their potential impact. Findings suggest that respondents recruited from these two sample sources are more alike than not, providing support for our blended sample approach. Differences in education and employment could impact study findings and require further investigation. Recruiting a representative national sample of a low prevalence population can be challenging and costly, making it critical to explore the impact of sample source on participant population and study outcomes.

Using Non-Probability Sampling Techniques to Track Seasonal Flu Activity

Zachary H. Lewis, *Ipsos Public Affairs*

Mary Choi, *Ipsos Public Affairs*

Influenza presents a major public health challenge and can cause serious illness and death. The Centers for Disease Control and Prevention (CDC) recommends vaccination for everyone age six months and older, although recent surveys indicate a vaccination prevalence of 58.9% among children and 42.2% among adults. The CDC monitors vaccination through a number of surveys, including the Behavioral Risk Factors Surveillance System, the National Immunization Survey, and online surveys. It tracks the incidence of seasonal flu through physician reports. This paper examines the possible use of a daily online omnibus survey that relies on opt-in panels and river sampling to produce weekly estimates of flu vaccination and flu incidence from December 2014 through March 2015. The results from Ipsos' daily online omnibus survey will be compared directly to prevalence and inoculation data reported by CDC. Comparability in

AAPOR 70th Annual Conference

terms of consistency, reliability and variability for main results and key population subgroups over the influenza season is explored.

The AARP National and State Online Fraud Survey

Jennifer Sauer, *AARP*

Joanne Binette, *AARP*

AARP commissioned the GfK Group to conduct a survey to better understand behaviors that may increase a person's risk of becoming an online fraud victim; what proportion of individuals nationally, and in certain states, may be at risk of becoming online fraud victims; and what steps are Americans taking to protect themselves? The GfK Group (GfK, formerly Knowledge Networks) fielded this survey on behalf of AARP. 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® as well as a supplement panel for some state-specific targets. A total of 11,741 surveys were completed from November 23, 2013 through December 30, 2013. While 8,150 were from KnowledgePanel® (KP), 3,591 were from an off-panel sample to supplement the state oversamples to ensure a minimum of at least 800 completions per state with the exception of South Dakota. The KnowledgePanel® completed 1,539 national sample surveys. This national survey has a completion rate of 51.3 percent and a margin of error of 3.1 percent. In addition to a state oversample in SD, other states that were oversampled are: AR, WA, TX, OH, FL, NH, NY, ID, AZ, GA, IN. To be able to generalize to the population, this study used unique and complex weighting scheme. Calibration weights blend off panel and KP responses and adjust for survey non-response and non-coverage or under- and over-sampling resulting from study specific design. Demographic distributions for the target population from the most recent CPS are used as benchmarks in this adjustment. All KP respondents were first weighted to these benchmarks. Weights were trimmed and scaled to all eligible KP respondents. KP and off-panel eligible respondents were then combined and weighted to the benchmarks of all eligible KP respondents.

Korean Occupational Prestige Scale: From KGSS(Korean General Social Survey)

Hong Joon Yoo, *Dept. of Sociology SKKU(Sungkyunkwan University)*

Shin Jin, *Graduate Student Dept. of Sociology SKKU*

* Supported by 2012 SSK(NRF-2012S1A3A2033331) The measurement of occupational prestige has been conducted in various ways. In order to find the best method to measure Koreans' perceptions for occupational prestige, four methods of measurement were compared. Those are; Type 1(a 5-point scale), Type 2(a 11-point scale), Type 3 (granting 100 points), and Type 4(ranking jobs in order). The standard deviation of Type 1 measure is the biggest while those of Type 2 and Type 3 are almost similar. The Type 4 showed the smallest. The correlation analysis among the four different types of measurement showed very high correlation coefficient between Type 2 and Type 3 measures. The respondents showed difficulty in their answer in sequence of Type4, Type3, Type2, and Type1. Overall, we concluded that Type 2 would be the best scale to measure occupational prestige score in Korean General Social Survey(KGSS) considering the time required, level of education, and age of respondents. After constructing scale, we upload a module in 2013 KGSS. And we analyzed Korean occupational prestige scores measured for one hundred selected jobs by using 2013 KGSS. The contents of our posters will include research purpose, previous studies, research model, results, and conclusion.

AAPOR 70th Annual Conference

A Neighborhood-Based Approach to Understanding Health and Healthy Living in Urban Communities

Katelyn Duffy, *Westat*

Denise St. Clair, *Westat*

Vasudha Narayanan, *Westat*

Ismael Flores Cervantes, *Westat*

Westat conducted an exploratory study under contract with the San Francisco Department of Public Health (SFPDH). The project included both quantitative and qualitative data collection that aimed to better understand general health and healthy living in San Francisco, CA, and inform larger studies in the future. The quantitative data collection consisted of a 4-page mail survey to a sample of households across all 11 districts of the city. Then, to add context and help fully understand the unique situation of the more complex districts in the city that are home to key subpopulations of interest, including Spanish and Cantonese-speaking residents, low-income residents, and other minority residents, focus groups and informant interviews were conducted. In this way, the quantitative work provided breadth, allowing an exploration of the topics of interest across all the districts of the city. And, the qualitative portion provided depth, allowing a better understanding of the primary issues at hand for the districts most in need. Together, data collection provided a wealth of information to help best inform future research that is statistically significant and relevant beyond the immediate study population. A primary finding was that to truly address the needs of the city's residents it is necessary to evaluate the city starting at the neighborhood, not district, level. Using a new approach looking at healthy living at a neighborhood level allows the city to best understand how diverse neighborhoods differ and how they are similar and how resources and attention can best be focused at a more micro level. This paper looks at lessons learned in San Francisco and how a neighborhood approach can be implemented in a resource-effective way.

How to Code School Names More Efficiently: Common Sense, Scripting and a Novel SAS Application

Akbar Akbari Esfahani, *University of California Los Angeles Center for Health Policy Research*

Matt Jans, *University of California Los Angeles Center for Health Policy Research*

Ninez Ponce, *University of California Los Angeles Center for Health Policy Research*

Coding open-ended text strings collected in survey interviews can be extremely time-intensive. We significantly reduced the processing time and human intervention required to clean and link school names recorded in the California Health Interview Survey (CHIS) to a list of school names provided by the California Department of Education (DOE). The coding process began by verifying the DOE list using GIS geocoding functions which had clerical errors in about 10% of the dataset, and included standardizing school names to maximize DOE list link potential as much as possible. Linking cleaned respondent-provided names to the DOE list involved four steps that were carried out in sequence. Each linking step included cases that remained unlinked after the previous steps. Those remaining unlinked after step four were processed with the SAS SOUNDINDEX function (Fan, SUGI 29; Knuth, 1973; Odell & Russell, 1918, 1922). SOUNDINDEX is a genealogical research tool that creates codes that will be linkable even if the original names were spelled slightly differently. The four link steps were then conducted on the SOUNDINDEX-coded data. Our test included 2920 cases, 705 of which remained after the first four-step pass and 138 after the four-step pass on the SOUNDINDEX coded names. Automation reduced overall coding duration from 4 weeks (140 actual work hours) with a human coder to 1

AAPOR 70th Annual Conference

hour of computer time plus an additional 8 hours for manual upcoding of remaining unmatched cases. The one-time start-up cost was about two working days for an experienced and proficient programmer. Knuth, D.E. (1973), *The Art of Computer Programming*, Volume 3. Sorting and Searching, Reading, MA: Addison-Wesley. Margaret K. Odell and Robert C. Russel (US Patents 1261167 (1918) and 1435663 (1922)). Fan, Z., (2004), *Matching Character Variables by Sound: A closer look at SOUNDEX function and Sounds-Like Operator*, Rockville, MD: Westat.

The Influence of Question Order on Social Desirability Bias in a Telephone Survey

Florian Keusch, *University of Mannheim*

Ting Yan, *Westat*

David L. Vannette, *Stanford University*

James Lepkowski, *University of Michigan*

Feeling thermometer questions allow respondents to express their attitudes toward a person, group, or issue on a numeric rating scale (usually from 0-100). Feeling thermometers are widely used in surveys and especially popular in surveys on political issues, but have been criticized for the lack of clear and uniform meanings of their many scale points that could be interpreted differently by respondents. Additionally, questionnaire designers might worry that respondents answer in a socially desirable way, if the feeling thermometer questions ask for ratings on sensitive groups of the society, such as gays and lesbians or illegal immigrants. In this study, we will examine the influence of the order in which target groups are presented to respondents on their ratings using the feeling thermometer. Specifically, we propose to assess the extent of social desirability bias in feeling thermometer questions as a function of the question order. Making use of an experiment in a CATI survey that asked respondents to rate their general attitude toward five groups of the society, respondents were either asked to rate their feelings toward these groups in an increasing order of sensitivity (Congressional Democrats, Congressional Republicans, politicians in general, illegal immigrants, gays and lesbians) or a decreasing order of sensitivity (gays and lesbians, illegal immigrants, politicians in general, Congressional Republicans, Congressional Democrats). We assess how the order of the questions affects the social desirability bias in the ratings of both sensitive and less sensitive groups using the feeling thermometer. The results of our study will inform questionnaire designers on how to best design feeling thermometer questions.

An Exploration of Survey Questions that Implicitly Categorize Respondents and Implications for Empirically Based Question Design Principles

Stephanie Willson, *National Center for Health Statistics*

Designing survey questions that capture the constructs intended to be measured is not a straightforward process. Indeed, the act of writing survey questions has often been likened more to art than science – an odd premise, given that the task of survey data collection is located squarely in the realm of statistical measurement. This paper takes a scientific approach to question design. Through an examination of findings from multiple cognitive interviewing studies, it will identify common interpretive patterns associated with disparate survey questions on health topics, summarize the connective theme among the patterns, and posit a general question design principle that is grounded in empirical evidence. Data are examined from reports in Q-Bank, an on-line repository of cognitive interviewing final reports. Survey topics include smoking behavior, cancer control, dental health, and alternative medicine. The author will show how survey questions often – and unintentionally – dichotomize respondents into categories relevant to the topic. For example, smoking questions may imply that the respondent is either a smoker or non-smoker. This may be problematic when the intended construct is

AAPOR 70th Annual Conference

behavior, not self-concept. Respondents who smoke but do not view themselves as “smokers” have a different question-response process than those who identify as smokers or non-smokers. Smoking is a straightforward example, but less intuitive examples also exist. This paper will demonstrate that different kinds of questions have the potential to (unintentionally) create dichotomous categorizations of self-concept, some of which do not exist a priori among respondents but, rather, emerge as an artifact of the question-response process. When respondents cannot place themselves in the categories assumed by a question, they struggle in choosing a response option, decline to answer, or interpret the question differently than intended. Therefore, when attempting to measure behavior, survey planners should avoid questions that imply the question is about self-concept.

Can An Importance Prompt Reduce Item Nonresponse For Demographic Items Across Web and Mail Modes?

Glenn D. Israel, *University of Florida*

Conventional practice usually places demographic items at the end of a questionnaire. The thinking behind this practice is that demographic items are less important than topically-salient items, so a higher risk for item nonresponse is tolerated. A recent study by Teclaw, Price and Osatuke (2012) turn this logic on its head and found that item response for demographic items at the beginning of a questionnaire was higher than for the same set of items at the end of the survey. This finding raises the question of whether there are other approaches to stimulating high item response rates for demographic questions. Other studies (e.g., Messer et al., 2012; Lesser et al., 2012) have found item response rates are higher for those responding by web than for mail. This study extends previous research by exploring how a statement about the importance of answering the demographic items at the end of a survey might mediate the relationship between mode and item response rate. Experimental data from a customer satisfaction survey of Cooperative Extension Service clients are used for the study. Overall, the aggregate item response rate was significantly higher on seven demographic items for web respondents than for mail respondents. Among web respondents, the importance prompt treatment had a statistically significant higher aggregate item response rate than the no prompt treatment. Conversely, the aggregate item response rate for mail respondents with the importance prompt was not significantly greater than that for mail respondents without the prompt. These results suggest that an importance prompt is a viable strategy for reducing item nonresponse of demographic items, at least among web respondents. One explanation for the results might be that using an importance prompt on single-question screens drew more attention to the prompt than having the prompt embedded among questions on the paper instrument.

The Problem of Double Confounding of Interviewer and Area Effects

Koen Beullens, *Centre for Sociological Research - KU Leuven*

Geert Loosveldt, *Centre for Sociological Research - KU Leuven*

The design effect due to interviewer clustering is usually formalized as $1+(m-1)\rho$, with m representing the average interviewer workload. The intra-class correlation ρ is the result of a two-by-two confounding effect: the interviewer versus area effects and the measurement versus nonresponse effects. Interviewers can have a different effect on the answers given by their respondents, but also respondents of different areas may have different characteristics. In addition, interviewers may have different response rates and/or attract different kinds of respondents, and areas may have different response rates. When interviewers and areas are very closely related, the specific effects of these four cells are hard to disentangle, making the

AAPOR 70th Annual Conference

estimation of these effects on statistics and their standard errors especially difficult to assess. This paper seeks to demonstrate that the traditional formula $1+(m-1)p$ is likely to underestimate these interviewer and area effects. This is due to (1) differences in workload among interviewers (so that interviewers need to be weighted to even out their specific measurement errors) and (2) the fact that areas should be given different weights to restore geographical representativeness, due to nonresponse. A third reason (3) is that since interviewers obtain different response rates and have different specialties, interviewers secure cooperation from people of variable response propensities, suggesting the need to apply weights on the interviewer level. It is crucial to see that these three interventions are very likely to be incompatible, particularly since each time the weights need to be applied to the same interviewer/area units. As a result, applying adjustments to solve problem (1) may deteriorate the situation in (2) and (3), and vice versa. The European Social Survey will be used to assess the design effect under various assumptions related to this two-by-two confounding of interviewer and area effects.

Saturday, May 16
2:15 pm. – 3:45 p.m.
Concurrent Session I

Mini-Conference: AAPOR Big Data Task Force Panel

Mini-Conference: AAPOR Big Data Task Force Panel

Frauke Kreuter, *JPSM*

Lilli Japec, *SCB*

Julia Lane, *American Institutes for Research*

Paul Biemer, *RTI*

Abe Usher, *humangeo*

Paul Decker, *Mathematica*

The American Association for Public Opinion Research (AAPOR) is a professional organization dedicated to advancing the study of “public opinion,” in a very wide definition including attitudes, norms, values, behaviors etc. As an organization AAPOR continually works to improve the methods and measures used to collect relevant data, to educate its members as well as policy makers, the media and the public at large to help them make better use of surveys and survey findings, and to inform them about new developments in the field. It is in this context that the AAPOR council saw the pressing need for a report that summarizes the discussion and challenges around Big Data for the AAPOR community. Also because AAPOR advocates the highest standards of ethical conduct of survey and opinion research, and promotes best practices in collecting, analyzing, and interpreting (survey) data, this task force report will have ethical issues in mind when it comes to the use of Big Data and tries to lay out the land for possible guidelines towards best practices and standards when using Big Data by themselves or in combination with survey data. Obviously methodology and procedures will vary with the data sources used. However, in the context of the use of Big Data this task force had these principles in mind when creating the report. The objectives of this report are in short: · to educate the AAPOR membership about big data · to describe the big data potential · to describe the big data challenges · to discuss possible solutions and research needs for big data challenges and · to provide guidance to researchers using big data.

Mode Effects

From Telephone to the Web: The Challenge of Mode of Interview Effects In Public Opinion Polls

Scott Keeter, *Pew Research Center*

Nancy Mathiowetz, *University of Wisconsin*

Kyley McGeeney, *Pew Research Center*

Ruth Igielnik, *Pew Research Center*

As survey research rapidly evolves, two notable trends are the growth of self-administered web surveys and mixed-mode studies. After decades in which most surveys were interviewer-administered, the new world of survey self-administration poses both opportunities and challenges. Among the challenges is how to handle long-term trend data when the mode of interview shifts, and how to minimize error in merging data in multi-mode studies. Mode-of-interview effects have been the subject of many studies, but rarely has a study examined mode effects across a wide ranging set of political and social measures. We build on past research with a comprehensive, large-scale mode-of-interview experiment that randomly assigned a nationally representative panel of respondents to telephone and web modes and administered a 75-question instrument to them. Respondents are drawn from the Pew Research Center's American Trends Panel, recruited in early 2014 in an RDD telephone survey. The experimental groups include 1,509 respondents who took the survey on the web, and 1,494 who completed a telephone survey. Each experimental group was weighted separately to national parameters for the general public. Non-internet users (12% of the panel) were interviewed by phone but are excluded from this analysis. This presentation will provide an overview of the size, direction and nature of mode differences across standard polling questions that varied in terms of topic, format, sensitivity and cognitive complexity, among other dimensions. The mean difference in topline results between web and phone across all questions was 6.4 points; median difference was 5.0 points. Differences ranged from 0 to 29 points. The analysis shows that topline comparisons can be misleading. Even on questions with small overall mode differences, sizeable mode effects may lurk beneath the surface, often appearing only among subgroups for which a particular question is sensitive or cognitively difficult.

An Evaluation of the Effect of Mode-Switching in Panel Surveys Using Recall Data

Nick Allum, *University of Essex*

Fred Conrad, *University of Michigan*

A key concern about the web survey data quality is difficulty garnering a probability sample because there are no good frames of email addresses for a general population. In a panel survey, it is possible to switch respondents to web mode after initial recruitment via face to face methods (F2F), thus mitigating the problem of sample selection and allowing the collection of rich information at lower marginal cost. However, web respondents generally seem more likely to take shortcuts than respondents in interviewer-administered modes (e.g., Heerwegh and Loosevelt, 2008). This may even be exacerbated by switching from F2F to web: by contrast to an interview, self-administration feels particularly "unsupervised" and, without an interviewer to motivate them to be conscientious, web respondents may take shortcuts and minimize their effort compared to their style of participation in previous F2F interviews. This raises the more general issue of whether it is possible to maintain the integrity of time-series in which there is a midstream mode switch (FTF to web). In this paper, we report results of an experiment in a

AAPOR 70th Annual Conference

panel survey (the UK Household Panel Survey Innovation Panel (UKHLSIP)) that compared the accuracy of past event recall, validated by responses at previous interview, comparing respondents who were switched to web mode with those that continued to be interviewed F2F. We also assess how asking for pre-commitment to careful answering might mitigate any potential loss of data quality resulting from switching to web. More generally, our results allow us to understand a little better how mode-switching interacts with cognitive processes underlying survey response to produce data of varying quality.

Characteristics of Web, Mail and Phone Responders to a Survey About the Health Insurance Marketplace

HarmoniJoie Noel, *American Institutes for Research*

Stacey Bielick, *American Institutes for Research*

Daniel Harwell, *American Institutes for Research*

Steven Garfinkel, *American Institutes for Research*

The Affordable Care Act (ACA) authorized the creation of Health Insurance Marketplaces (Marketplaces) to help individuals shop for, select, and enroll in private health plans beginning in October 2013. In order to measure consumer experiences with these Marketplaces, the Centers for Medicare & Medicaid Services (CMS) established the Health Insurance Marketplace Experience Survey. A mode experiment was embedded in the 2014 field test of this survey in order to determine the most cost-effective ways to conduct this survey in future years while maximizing data quality. Sampled individuals were assigned to five experimental groups: (1) telephone-only, (2) mail with telephone follow-up, (3) mail with FedEx follow-up, (4) Web-only, and (5) First Class Mail only. This paper will explore the potential for mode bias given that some demographic groups are more likely to respond to different modes (mail, phone, web) and whether those differences affect key survey estimates. Preliminary results suggest that consumers who are white, wealthier, more educated, and have had health insurance in the past year tend to complete the survey by web more often than other modes; whereas, consumers who are racial or ethnic minorities, less educated, and have not had health insurance in the past year are more likely to respond by mail or phone. Additionally, this paper will explore the potential for mode bias in follow-up modes to determine whether phone or FedEx follow-up modes bring in different subgroups of respondents compared to first class mail follow-up to decide whether higher cost modalities are worth it. Finally, we will explore whether these differences impact key outcomes, such as the overall rating of the Marketplace. Results from this paper will impact decisions about which mode or modes to use in subsequent data collections for the Marketplace Survey as well as other surveys of Marketplace users.

Scrub-A-Dub: Telephone Sampling Scrubbing and Working Rates

Landline Frame Changes From 2010 Through 2014 And Their Implications

Marilyn Wilkinson, *Abt SRBI*

Dennis Daly, *Abt SRBI*

Barbara Fernandez, *Abt SRBI*

Andrew Burkey, *Abt SRBI*

Whereas landline households remain easier to target geographically, more amenable to demographic targeting, and generally less expensive to interview compared with cell phone only households, the quality of the landline frame is declining. Our paper explores recent changes in

AAPOR 70th Annual Conference

landline frame quality by examining the Behavioral Risk Factor Surveillance System (BRFSS) landline frame pre-screened sample records for five states from 2010 through 2014. We also discuss the impact of these changes on completion rates and production costs for landline completes. BRFSS uses sample numbers that have been prescreened as working and belonging to households. Among the five states examined, we found a significant decrease in the proportion of the landline sample prescreened as households, from 47% in 2010 to only 39% in 2014. We also found significant shifts over time in groups of disposition codes (complete, eligible non-interview, unknown eligibility, not eligible) in the dialed sample of numbers prescreened as working household numbers. Between 2010 and 2014, the proportion of completed interviews from the prescreened sample decreased from 9% to 4%, the proportion of eligible non-interviews decreased from 4% to 2%, and the proportion of not eligible numbers increased from 66% to 74%. Accounting for these changes in the landline sample, production costs for achieving completes have increased steadily from 2010 to 2014. We will discuss the implications of our findings for conducting landline and dual frame telephone surveys.

Use of Phone Number Usability Indicators to Improve Calling Efficiency and Reduce Cost

Robin J. Gentry, *The Nielsen Company*

The working phone rate, the percent of assigned residential numbers in a sample, has dropped consistently over the past two decades. A sharp increase in demand for additional area codes and exchanges has been driven by the rise in cell phone ownership, VoIP usage, and cable companies providing phone service. As the working phone rate drops, additional effort is required to screen phone numbers and determine whether or not they are associated with a usable household—resulting in increased cost and cycle time. For years, many research companies have screened sample against lists of known businesses. However, more recently, two sample vendors have begun offering additional services to assess phone number usability. Survey Sampling International (SSI), offers a variable known as an activity indicator and Marketing Systems Group (MSG) provides a variable known as Cell-WINS, both of which can provide meaningful information about the likelihood that a number is working. Nielsen, a leading global information and measurement company, recently tested both of these indicators on a large syndicated survey of radio audience measurement. The indicator variables were appended after scrubbing against known business numbers and screening via MSG's ID Plus service (an attended screening process which removes additional business and non-working numbers). The phone numbers were then called up to 13 times before receiving a final disposition code. Next, the final disposition was compared to each indicator, independently, and to the cross-classification of the two indicator variables. Through this process several interesting and potentially actionable subgroups emerged. This presentation will focus on how these vendor appended usability indicators can be used to determine which numbers to attempt and the level of effort to apply to those attempted numbers. We will also demonstrate that applying these rules can result in decreased cost and improved efficiency when calling landline RDD sample.

Is That a Good Phone Number? Scrubbing Phone Numbers Appended to an ABS Sample

Lawnzetta T. Yancey, *The Nielsen Company*
David Malarek, *Marketing Systems Group*

In an effort to increase the efficiency of address based sampling, Nielsen investigated scrubbing phone numbers appended to the sample. For the TV diary sample, Nielsen uses a multi-mode

AAPOR 70th Annual Conference

design (phone/mail). Addresses matched with a phone number are recruited by phone while those without a phone number are recruited by mail. The phone numbers appended to our samples are from large consumer databases and the age of the data can vary depending on the source. On average, we call over 900,000 telephone numbers four times a year. We normally find that ~40% of the phone numbers are categorized as unusable (not in service, out of order, or no answers) after recruitment. To improve the effectiveness of calling and recruiting households for our surveys, we investigated the accuracy of two different technologies available from our vendor (Marketing System Group). The goal of the test was to see if one of two pre-screening methods could accurately identify unusable phone numbers in the sample. If the test shows that the screening process identified unusable numbers with a high level of accuracy, we would want to give those addresses an opportunity for recruitment by changing them to unmatched (not part of the test). This paper will review the results of a test conducted using phone numbers selected for our May 2014 TV diary survey.

Implications for Weighted Survey Estimates When Flagged-Inactive Cell Phones Are Excluded or Subsampled

Courtney Kennedy, *Abt SRBI*

Marci Schalk, *Abt SRBI*

Kyley McGeeney, *Pew Research Center*

Chintan Turakhia, *Abt SRBI*

Dean Williams, *Abt SRBI*

Leveraging activity codes has emerged as a promising approach for increasing the cost effectiveness of cell phone RDD samples. Provided by a sample vendor, activity codes categorize cell phone numbers as active, inactive, or unknown based on estimates of the working status of each number. Activity codes can be used to remove likely inactive cell phone numbers before the start of interviewing. This helps to reduce the amount of time interviewers spend dialing non-working numbers. That said, the approach is relatively new, carries a nontrivial cost per record for appending the code, and entails bias/variance trade-offs that should be carefully considered. To date, the research on activity codes has emphasized total exclusion of flagged-inactive numbers and focused on implications for unweighted full sample estimates. This is an incomplete picture of the ways in which activity codes can be used and the consequences for data quality. This paper builds on the existing literature in three ways: (1) we consider subsampling flagged-inactive cases as an alternative to full exclusion, (2) we evaluate the effects of using activity flags on weighted estimates, as opposed to unweighted estimates, and (3) we examine subgroup estimates (e.g., for young adults, rural residents, etc.), as opposed to just full sample estimates. The data come from a large (n=10,000) national overlapping dual frame RDD survey of U.S. adults conducted for the Pew Research Center in early 2014. Activity codes were appended to the cell sample but not used in the sample design, allowing us to evaluate their performance and simulate a design that under-sampled or excluded flagged-inactive cases. In this paper we report the major findings from that research. Results are presented for both full sample and subgroup estimates. This research provides survey designers more information about the effects of inactive exclusion or subsampling on final weighted estimates.

AAPOR 70th Annual Conference

Predicting Sample Performance in Telephone Surveys

Kristie M. Healey, *ICF International*

Joshua Brown, *ICF International*

Estimation of sample needs is a challenging part of conducting a telephone survey. Selecting too few telephone numbers can result in loss of precision if the target number of interviews is not met. Conversely, a project's budget may not be able to support the excess interviews if too many telephone numbers are selected. Splitting a study into smaller waves can allow for adjustments to be made based on past performance. However, the timing of survey fielding often requires selecting sample for the next wave before the current wave is completed. This paper presents a simple model for estimating the number of interviews that will result from a telephone survey in progress. Using previously completed wave from the same study, telephone numbers are assigned a probability of eventual completion at each attempt based on the call history up to that point. The probabilities are then applied to records in the current data collection wave with similar call histories. This model was used to estimate monthly interview yields for a dual-frame national survey measuring smoking and related behaviors of adult U.S. residents. Monthly predictions were within 8 percent of the final totals for landline studies and within 15 percent of the final totals for cell studies. The model is evaluated to determine the point where optimal accuracy occurs. Additionally, we examine the model's performance in a variety of study types, including those where only a subset of adults are eligible.

Media Influence on Public Opinion

Do Polls Drive the News or Does News Drive the Polls? A Meta-Analysis of Polling Questions and Big News Stories

Kristie Jo Taylor, *University of Connecticut*

Critics of public opinion polling in the U.S. often note the problematic relationship between mass media outlets and survey research. One major criticism is that media outlets rely on polls to frame their news stories, thereby obscuring the real issues at hand. Others have put forth the theory that news media choose to focus more on issues that divide the public in order to provide news content that boosts ratings but does not arm the public with an in-depth understanding of the issues, an important aspect of any healthy democracy. The assumption is that media use polls to find the contentious issues to report; yet no empirical research has been conducted to support or refute this claim. Using data from the Roper Center for Public Opinion Research archives and the Pew Research Center's Project for Excellence in Journalism News Coverage Indices from 2010 and 2011, this paper examines the relationship between media coverage of issues and polling on those issues. I ask: Do poll questions on specific issues appear before or after peaks in media coverage on those issues? Do polls sponsored by media outlets differ from other types of polls in the timing of these questions? Results of the meta-analysis show a more complex relationship between public opinion polls and the media than critics acknowledge.

AAPOR 70th Annual Conference

What's Wrong With the News? Perceptions of News Coverage Among African Americans and Hispanics

Malaena Benz, *The Associated Press-NORC Center for Public Affairs Research*

Nicole Willcoxon, *The Associated Press-NORC Center for Public Affairs Research*

Emily Alvarez, *The Associated Press-NORC Center for Public Affairs Research*

Tom Rosenstiel, *The American Press Institute*

Trevor Thompson, *The Associated Press-NORC Center for Public Affairs Research*

Jennifer Agiesta, *The Associated Press*

While many forecasters of the digital future have predicted a digital divide, in which people of color would be left behind in the use of technology, a new study about news consumption finds that this is not playing out as anticipated. The Media Insight Project, a collaboration between the American Press Institute and the Associated Press-NORC Center for Public Affairs Research, addressed these issues using a nationally representative random digit dial survey with African American and Hispanic oversamples. The results of this work show that these two minority groups are using digital technology for news at similar rates as the American population overall. Yet these minority communities do not believe that the growth of web and mobile media has fulfilled the promise of more coverage, or more accurate coverage, of their underserved racial and ethnic communities. The study uses a unique approach to measure news consumption habits and compares the news preferences and behaviors of African American and Hispanic adults to adults nationwide, as well as to non-Hispanic whites alone. Even though the classic concerns about a digital divide based on connectedness do not fully describe the landscape, the survey finds some important differences among racial and ethnic groups when it comes to news consumption after controlling for age, education, income, and other demographic and social factors. Those differences mainly have to do with where people go for news and the topics and issues they follow in the news. The findings are based on a nationally representative telephone survey of 1,492 adults, including an oversample of 358 Hispanic adults and 318 African American adults. Interviews were conducted in English and Spanish, and on landline and cell phones. Data were collected by NORC from January 9 through February 16, 2014.

Political Conspiracies: Who Believes What and Why?

Jennifer Cassino, *Fairleigh Dickinson University PublicMind Poll*

Krista Jenkins, *Fairleigh Dickinson University PublicMind Poll*

Peter Woolley, *Fairleigh Dickinson University PublicMind Poll*

Despite conventional wisdom regarding the inherent rationality of public opinion, beliefs in political conspiracies are a fixture in the American psyche. Our research has found sizable numbers who remain convinced that, among other things, President Obama is hiding information about his background and early life (36%), President Bush knew about the 9/11 attacks before they occurred (25%), and that both the Democrats (19%) and Republicans (23%) have engaged in coordinated voter fraud campaigns in order to win presidential elections. In short, those who dismiss evidence to the contrary and believe in conspiracies can be found in numbers that defy consideration of them as "just the fringe." Having already measured beliefs in political conspiracies across a variety of domains, our proposal now rests on the expectation that knowledge and media exposure are confounding variables when it comes to understanding those who do and do not believe in conspiracies. Preliminary findings from our past research attest to the role that political knowledge plays, as we found in a 2012 nationwide survey. Specifically, the more informed you are, the less likely you are to endorse political conspiracies. Education also tends to reduce belief in conspiracies. However, left unexamined is the role that

AAPOR 70th Annual Conference

the media plays. Is news consumption related to if you believe, what you believe, and how strongly you believe in political conspiracies? We expect to find patterns of relationships that broadly adhere to liberal and conservative news outlets, as well as consumption of print versus broadcast news sources, and will examine these and other questions related to beliefs in political conspiracies through a national telephone survey we will field in December 2014. We will expand our focus to examine the precursors to conspiratorial beliefs, as well as cast a wider net by measuring beliefs in a host of other political conspiracies.

Real-World Agenda-Setting: Linking Different Types of Voters to Media Reporting on Political Events

Dan Bytzek, *University of Koblenz-Landau*

Ina E. Bieber, *Goethe-University Frankfurt*

Sigrid Rossteutscher, *Goethe-University Frankfurt*

Philipp Scherer, *Goethe-University Frankfurt*

Although there is ample research on media agenda-setting, it suffers from two shortcomings: First, the link between the media agenda (a macro phenomenon) and voters (at the micro level) is often indirect since research is either experimental or uses aggregated public opinion as dependent variable. Second, when putting the individual voter into focus, it remains unclear if different types of voters also react differently to media reporting, e.g. with regard to media usage or party identification. This paper thus uses multi-level analysis with media reporting on the macro and individuals' opinions on the most important problem on the micro level by combining several online surveys and media content analyses conducted by the German Longitudinal Election Study. This enables us to study both real-world agenda-setting effects and interactions between media reporting and voter characteristics, hence to contribute an important piece of the puzzle to media agenda-setting research.

Transparency, Survey Literacy and Motivated Reasoning in the Public Interpretation of Poll Results

Evelyn Kuru, *University of Michigan - Ann Arbor*

Michael Traugott, *University of Michigan - Ann Arbor*

Josh Pasek, *University of Michigan - Ann Arbor*

When do citizens listen to poll results? Although a large body of literature documents the impact of public opinion polls on opinion formation, little research has examined the processes of how this occurs. In particular, recent high-profile claims of "polling bias" and attempts to "unskew the polls" highlight variation in individuals' willingness to accept the results of even the highest quality surveys. Using a survey experiment, the current study explores three variables that might influence the public's willingness to accept poll findings: technical knowledge about surveys, the transparency of survey results, and the correspondence between survey results and preexisting issue positions. In the context of a news article, we expose a broad sample of Americans to polling information on three national issues – abortion, marijuana legalization, and American intervention in the Mideast – and assess how credible they find the survey results. We also vary whether the news sorties include the sort of information proposed by the AAPOR transparency initiative. Results suggest that individuals are more likely to believe surveys that correspond with their pre-existing attitudes and that the willingness to accept survey results is itself moderated by the transparency of the poll as well as respondents' basic poll literacy levels. This suggests that motivated reasoning in acceptance of news surveys may be offset by

transparency and polling knowledge. We discuss the implications for AAPOR's transparency initiative as well as for those reporting survey data.

Cultural Consideration of Interviewing in Africa & the Middle East

How Does Interviewer Social Status Affect Self-Reported Attitudes About Democracy? Evidence from 20 African Countries

Zeina Q. Lau, *RTI International*

Interviewers' demographic characteristics can influence how respondents answer survey questions, potentially biasing survey estimates. However, little is known about interviewer effects in developing countries—particularly about the role of interviewer socioeconomic status. Understanding these effects specifically in developing countries is important because interviewers in these settings typically have higher socioeconomic status than respondents. Higher status interviewers may elicit more over-reporting of socially desirable behaviors (e.g., voting) and more underreporting of socially undesirable attitudes (e.g., political distrust, pessimism about democracy.) This study investigates how interviewer social status affects self-reported attitudes about democracy in 20 African countries. Data from the 2008 Afrobarometer Surveys, in which 810 interviewers administered standardized surveys to 27,713 respondents across 20 countries throughout Africa, were analyzed. These surveys contain rich data about interviewer characteristics and multi-dimensional measures of respondents' political attitudes. In this paper, I use multi-level models to investigate the effects of interviewer social status (education, age, urban residence, language, ethnic group) on respondent attitudes towards democratic governance, political ideology, and political actors. Analysis suggests that interviewer social status has substantial effects on reports of political engagement and attitudes about democracy. For example, higher interviewer education is associated with respondents expressing more engagement with public affairs, greater fear of political intimidation, and more disbelief that their country meets democratic ideals. Similarly, interviewers from urban areas also elicited greater distrust of political system and more pessimistic attitudes about democratic governance. Additional analysis examines the effects of various combinations of interviewer and respondent characteristics, and also explores heterogeneity in interviewer effects across countries. This paper discusses the implications of these interviewer effects for designing and implementing field surveys in developing countries.

The Implications of Gender-Matching on Survey Research in Egypt

Charles M. Beaird, *D3 Systems, Inc.*

Samuel Solomon, *D3 Systems, Inc.*

Considering the social and political instability in Egypt following the Arab Spring, face-to-face research proves a formidable challenge. While dual-mode telephone sampling techniques have been extensively explored and developed in the Western context, data informing sampling plans in countries like Egypt are scarce. In 2013, D3 Systems, Inc. designed and implemented its own national face-to-face multi-stage probability survey to study both landline and mobile telephone ownership and telephony habits of Egyptian nationals. This paper utilizes the data to explore the role interviewer gender potentially has on participation rates and responses, in order to further inform and improve dual-mode telephone research methodologies in Egypt. Through a cross-comparison of the data based on interviewer and participant gender, the authors investigate

AAPOR 70th Annual Conference

how gender-matching affects both the likelihood of respondent participation and the nature of the responses provided.

Research in Afghanistan: Strategies for Overcoming Methodological Challenges

Jessica Besmel, *Northern Arizona University*

Frederic Solop, *Northern Arizona University*

Much of the mainstream literature on survey research is written from the perspective of researchers conducting work within western, industrial societies. At the same time, we know that similar methodologies and techniques are being adapted for use in to societies and cultures throughout the world. Pew Research Global Attitudes Project is but one example of this work. Afghanistan has recently been the subject of significant social science inquiry, and the source of great challenge to western, social science researchers. In particular, Afghanistan, like many countries, raises important questions about sampling, culture, and quality control. In the area of sampling, Afghanistan has not conducted an official census since 1979; therefore, the true population of the country is not known. Culturally, women do not participate in most surveys. Thus, surveys tend only to reflect of male opinion. In terms of quality control, security and difficult geographic terrain raise the costs of reaching certain segments of the population and verifying the quality of participation in these areas. This paper investigates how research companies are struggling to overcome these challenges in Afghanistan today. Specifically, this paper presents a case study of Afghanistan Research Services's (a Division of Afghanistan Holding Group) Mortgage Market Assessment, conducted in five major cities of Afghanistan. After outlining the goals of the research project and the specific challenges encountered, this paper explores strategies employed to overcome these challenges. This paper provides interesting insight into research methodologies used in Afghanistan today, and offers lessons for ambitious researchers interested in adapting standard survey techniques to future work in third world countries.

Do The Media Construct Gender ? A Comparative Study of Traditional and New Media's Role in Constructing Egyptian Youth Gender's Perceptions

Parwez Ahmad El Badaoui, *Cairo University*

This study aims to explore the impact of traditional and new Egyptian media in the perception of Egyptians university students of gender roles and attributes. To conduct this study, a questionnaire was distributed to a sample of 200 students in both public and private universities. The results showed a relationship between the frequency of exposure to traditional media and certain gender attributes and roles (often negative) related specifically to women. The predicted effect of exposure to both traditional & new media objectified gender models and actual gender-perceptions among Mass media's students were demonstrated in the current study. As expected there was a connection between the exposure's frequency to Egyptian drama and certain attributes and roles related to women in particular. The present results also showed that highly usage of new media was associated with forming positive attitudes and perceptions about women compared to men in terms of social, political and economic aspects. The study results also indicated a predominance of traditional roles, "inherited and customary" of men and women, which is associated with their biological and cultural nature in the Egyptian society, where men care for their jobs, women are busy with home affairs, raising children, shopping and sit down with friends of the same sex. In this context, it is recommended to employ the Internet in a right way to change and strengthen the positive perceptions of gender's roles & attributes in the Arab and Islamic societies. This could be best achieved through renewing the speech

associated with each of both men and women, with help of the developments that exist on the ground.

Panel Survey Refinements

Let Me Tell You What You Told Me: Dependent Interviewing in Establishment Surveys

Souraya Ridolfo, *National Agricultural Statistics Service*

Jennifer Edgar, *Bureau of Labor Statistics*

In panel surveys, respondents are asked to report the same information several times. Often, respondents are instructed to provide the same information for a different reference period or to update their previous answers. Survey designers are faced with the choice of whether or not to provide respondents with their previously reported data (PRD). On one hand, PRD may reduce respondent burden and improve data quality by giving respondents an anchor on which to base their current reports. Rather than starting from scratch, respondents can start with their prior reports and determine what changes or adjustments need to be made. On the other hand though, the risk of respondent satisficing may lead to lower quality data, particularly when the survey is attempting to measure change from one wave to another. Previous research on household survey respondents suggests that the impact of giving respondents their prior information can have differential effects, depending on the stability, saliency and complexity of the question topics (Sala & Lynn 2004; Jackle 2005; Jackle 2006). It can be difficult to generalize results from household surveys to establishment surveys however, and the limited research on establishment surveys was inconclusive (Tran & Gerling 2009; McCarthy 1995; Bailey 1994). Especially in interviewer-administered surveys, currently there is not much known about the impact of PRD on the response process in establishment surveys. This study seeks to understand, through cognitive interviews with establishments, how establishment respondents incorporate PRD when formulating their survey responses. Using the Occupational Employment Survey (OES) as a case study, we explore how respondents react to their PRD, and how, if at all, they use it when arriving at their answer to the current survey.

Flexible Stratification

Heather Josserand, *The Nielsen Company*

William Waldron, *The Nielsen Company*

When managing a household panel, there is a natural turnover and an effective recruiting process is necessary. The recruiting can be complex, especially when the panel needs to match various characteristics (demographics, geography, etc.). When the number of characteristics is relatively small, the recruiting can easily be achieved by a stratified sample design. When a lot of variable needs be taken into account, an alternative solution would be to use a balanced sample design but it can be a complex procedure. How could we keep advantage of the simplicity of stratification without its draw back when there are too many strata? The solution we are considering here is strata pooling. The goal is to reduce the number of strata in order to make it easier to manage and keeping the most flexibility possible. Each household belongs to an atomic stratum having various characteristics which are categorical variables. A stratum has also its own response rate. The problem is to draw a sample which will fit as close as possible different targets. This multivariate sampling allocation problem can be approximately solved by a Lagrange multiplier for one set of aggregated atomic strata. Using an ascending clustering

AAPOR 70th Annual Conference

algorithm on the strata and measure the smallest sample size matching the targets provides a very interesting solution. Number of aggregated atomic strata and expected sample size can be visualized simultaneously. We will demonstrate how the technique performed on 48 markets of Nielsen Audio. Hence, we will provide a straightforward method for survey administrators to determine how many strata they should keep to achieve a flexible strata allocation.

Using Paradata to Predict Case Completion Outcomes on the General Social Survey (GSS)

Etienne Daquilanea, *NORC at the University of Chicago*

The General Social Survey (GSS) provides a suitable environment in which to explore the use of paradata for predicting respondent cooperation. The GSS is a survey sponsored primarily by the National Science Foundation and conducted biennially by NORC at the University of Chicago to collect data on the attitudes, experiences, and demographic characteristics of residents throughout the United States. The 2014 round of the study in particular included a panel sample of respondents who completed the GSS interview in both 2010 and 2012. After completing the GSS interview, each interviewer rates the respondent's level of cooperation and level of comprehension. Ratings from previous rounds, coupled with paradata about dates of completion as well as numbers of contact attempts, refusals, and appointments broken before case completion, may be analyzed against the completion status of each case at the end of the subsequent round to create a model for predicting likelihood of case completion. In 2012, we analyzed the correlations between these paradata and completion outcomes and found that respondents with higher cooperation and comprehension levels as well as lower numbers of refusals in 2010 were significantly more likely to complete the interview in 2012. We are expanding this research for the 2014 round by combining paradata from both 2010 and 2012 to uncover how paradata would inform case completion outcomes into the third round of the panel. We will include demographic data in our analysis for the GSS 2014 round. These findings will add to the body of knowledge on using panel respondent paradata from current and prior rounds of a study to predict likelihood of subsequent case completion as well as further work towards creating a model that would use multiple previous rounds of paradata to predict completion in subsequent rounds of a study.

Generating Synthetic Longitudinal Data for Radio Reach

Jodie Sotiris, *The Nielsen Company*

Jiaquan Fan, *The Nielsen Company*

Etienne Josserand, *The Nielsen Company*

William Waldron, *The Nielsen Company*

When data is collected on multiple individuals over a period of time, it is sometimes the case that the resulting sample does not capture the information needed to perform a desired analysis. In particular, individuals in a longitudinal sample may not be followed for a suitable period of time to collect the necessary information. At an extreme level, only cross-sectional data may be available. One potential solution is to generate synthetic data for each sampled individual in a fashion that preserves, as much as possible, both the respondent's longitudinal distribution (when available) and the overall distribution measured over the full survey. Nielsen Audio records radio listening across the United States through a combination of electronically-measured radio panels in the larger markets, and with traditional paper "diary" surveys in the remaining smaller markets. The data is used by advertisers to estimate the percentage of a target population that was exposed to a commercial over the course of an advertising campaign. In metered markets, longitudinal data is collected over extended time periods, but we must still

AAPOR 70th Annual Conference

account for high non-response. The diary markets are even more problematic, since Nielsen does not follow these households beyond a week. This makes it impossible to know what a respondent had listened to prior to or after their measurement week. This paper discusses the challenges and successes of an attempt to generate synthetic longitudinal data for each diary respondent in a fashion that retains, to high degree, both the respondent's listening distribution and the overall listening distribution measured at the market level. Such an approach can yield several advantages, and may potentially be used in broader survey applications.

Attrition in Digital Ratings Panels

Ekaterina L. Oltman, *The Nielsen Company*

Attrition is a concern for all types of panel studies. Attrition is costly because ongoing efforts have to be made to recruit new members into the panel, and it can also lead to nonresponse bias if reasons for leaving the panel are related to behaviors measured in the panel. This paper investigates attrition in digital ratings panels where respondents install a meter on their device that measures app, web, and other usage. Unlike in traditional survey panels, once the respondent joins the panel, and the meter is installed, there is no further action required by the panelist. Opting out of the panel requires deliberate action such as communicating the desire to opt-out or actively uninstalling the meter. Kaplan-Meier survival curves and Cox proportional hazards models are used to examine the effect of demographic and device usage characteristics on risk of panel attrition. The single largest effect on panel attrition was due to recruitment mode. Panelists recruited as part of a probability sample using face to face recruitment had much lower rates of attrition than panelists recruited using online non-probability methods. Results also suggest gender, age, race, income, and education play a role in attrition. Men, younger respondents, low income respondents, and respondents without a college education were all more likely to leave the panels than other groups. Operating system (iOS versus Android), marital status, ethnicity, and employment status had no significant effects on survival probabilities. Behavior, measured as number of sessions and duration of sessions on the device did not affect panel tenure.

Challenges of Building + Maintaining Consistent Panels in Emerging Markets

Tim Richman, *GeoPoll*

King Beach, *GeoPoll*

Matthew Harber, *GeoPoll*

Roxana Elliott, *GeoPoll*

Research can be incredibly difficult in emerging markets; lack of widespread internet and electricity means many surveys are conducted by researchers going door-to-door with paper surveys, and respondents are often hard to track down after taking an initial survey. Finding respondents who fit certain specific criteria is equally frustrating, but many large companies and NGOs are looking for panel-based research to answer important questions on brand preference and consumer goods, or reach aid beneficiaries. This paper will look at a new method of remote data collection, the mobile phone, and explore how to build consistent panels of respondents through text-message surveys. While there are several challenges associated with data collection through mobile phones, including data validation, it can be a useful tool for building and conducting panel-based surveys in emerging markets. This paper includes key lessons learned when building and maintaining panels, including findings on respondent drop-off, survey length, survey frequency, and incentive amount. We will also discuss how to validate mobile data, ensure questions are worded clearly and respondents' answers are consistent over time, and when to remove respondents from a panel. Several months of testing with panel-based

AAPOR 70th Annual Conference

surveys has led to findings on how often to send surveys to a panel and in what order, and what techniques should be used for valid responses from a mobile population. Question and answer randomization and a reverse-ordering methodology have also improved data quality. Finally, we will go over unforeseen challenges when creating loyal panels in emerging markets, such as Mobile Network Operator outages.

Panel: Interviewer-Respondent Interactions in a Total Survey Error Framework

Panel: Interviewer-Respondent Interactions in a Total Survey Error Framework

Brady T. West, *University of Michigan*

Frederick G. Conrad, *University of Michigan*

Frauke Kreuter, *Joint Program in Survey Methodology/IAB*

Felicitas Mittereder, *University of Michigan*

Interviewers have long been known to affect the quality of survey data. However, the mechanisms by which interviewer effects occur are not well understood. Although traditional models for interviewers have treated interviewer effects as arising out of interviewers' influence on survey measurement (e.g., Kish 1962), recent research has shown that interviewer effects result from both differential nonresponse biases across interviewers and differential measurement effects (e.g., West and Olson 2010). One hypothesis is that interactions between respondents and interviewers lead to effects of interviewers on survey estimates; these interactions could occur at either the recruitment or measurement stage. This panel has five papers aimed at understanding interviewer-respondent interactions in surveys. West, Conrad, Kreuter and Mittereder empirically compare interviewer variance effects in standardized and conversational interviewing methods, each a package of interviewer behaviors which may differentially affect survey responses at either the nonresponse or measurement error stage. Schaeffer and colleagues examine interactions in survey recruitment calls, focusing on effects that interviewers have on unit nonresponse, and in particular, refusal conversion. Olson and Smyth evaluate changes in interviewer and respondent behaviors over the course of a survey field period, attempting to understand why interview length shortens as interviewers gain experience. Holbrook, et al. investigate how interviewer behaviors and question characteristics affect respondent behaviors. Finally, Smyth and Olson focus in on behaviors in a particular subset of questions – open-ended questions – and how these behaviors are related to the accuracy of recording answers. These five papers contribute greatly to our knowledge of the effects of interviewers in sample surveys, as well as data quality more generally. The findings of these five papers speak broadly to standardization of behaviors at recruitment and measurement, with implications for interviewer training and survey practice.

Comparing the Interviewer Variance Introduced by Standardized and Conversational Interviewing

Brady T. West, *University of Michigan*

Frederick G. Conrad, *University of Michigan*

Frauke Kreuter, *Joint Program in Survey Methodology/IAB*

Felicitas Mittereder, *University of Michigan*

Standardized Interviewing (SI) requires survey interviewers to read questions as worded and provide only non-directive probes in response to questions from respondents. While many major

AAPOR 70th Annual Conference

surveys use SI in an effort to minimize the effects of interviewers on data quality, the existing literature shows that between-interviewer variance in key survey statistics still arises despite the assignment of random subsamples to interviewers. Because this type of between-interviewer variance affects the precision of a survey estimate just like sample size, it has direct cost implications when designing a survey. Survey methodologists suspect that despite proper training in SI, interviewers may still diverge from scripts (even though they are not trained to) because additional explanation is often requested by survey respondents. Conversational Interviewing (CI) is known to handle clarification requests in a more effective manner: Interviewers are trained to read questions as worded, initially, and then say whatever is required to help respondents understand the questions. Despite literature demonstrating that CI produces noticeable decreases in the measurement error bias of survey estimates, survey researchers (and governmental agencies in particular) have been hesitant to employ it in practice, in part because of increased questionnaire administration time but also due to the fear of increased interviewer variance. We will present initial 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 use either CI or SI. This study featured an area sample and an interpenetrated sample design, enabling estimation of interviewer variance, and the presence of administrative information on the sampling frame also enables estimation of interviewer variance in measurement bias. We will employ analytic techniques designed to compare the interviewer components of variance between the SI and CI groups of interviewers.

Refusal Conversions Across Calls: Interviewer's Actions in Initial Calls and Their Consequences

Nora Cate Schaeffer, *University of Wisconsin-Madison*

Dana Garbarski, *Loyola University Chicago*

Jennifer Dykema, *University of Wisconsin-Madison*

Douglas W. Maynard, *University of Wisconsin-Madison*

Bo Hee Min, *University of Wisconsin-Madison*

Ellen Dinsmore, *University of Wisconsin-Madison*

We describe the interaction leading to initial refusals for sample members who will subsequently be converted and those who will not to determine whether actions of interviewers in the initial contact have different consequences for these two groups. Our goal is to identify actions of the interviewer in the initial contact that prepare the sample member for future conversion—or possibly harden the sample member in their refusal. Recent studies have predicted subsequent participation using householder concerns and other features of interaction for face-to-face interviews (e.g., Bates et al. 2008; Durrant & D'Arrigo 2014). We complement those studies by describing, for telephone interviews, the actions of interviewers during refusals that are subsequently converted and those that are not. We examine how the sample member declines (e.g., “not interested”) and what subsequent actions the declination makes possible. Other actions include topical tailoring (Couper and Groves 2002); other forms of responsiveness (e.g., agreeing with the sample member); which interviewer actions elicit a repetition of the declination; proposals by the interviewer (e.g., that the timing of the call is bad); and various persuasion strategies. In addition, we document how the exit from the call is structured and its association with subsequent participation. We use recordings from the 2004 wave of the Wisconsin Longitudinal Study, a panel study of Wisconsin high school graduates in 1957 (overall response rate approximately 80 percent). Actions were identified and coded based on a content and conversation analysis. From an analytic sample for a case control study of 257

AAPOR 70th Annual Conference

paired declinations and acceptances matched on the sample members' propensity to participate in the survey, we compare features of the interaction across the 76 pairs in which the refusing sample member was subsequently converted in a later call and the pairs in which the refusing sample member was not converted.

Why Do Interviewers Speed Up? An Examination of Changes in Interviewer Behaviors Over the Course of the Survey Field Period

Kristen Olson, *University of Nebraska-Lincoln*

Jolene D. Smyth, *University of Nebraska-Lincoln*

Researchers have observed that interviewers affect the length of a survey interview over the course of a field period by systematically speeding up. How interviewers speed up is unknown. One hypothesis is that certain interviewer behaviors are omitted or shortened over the course of a field period. A second hypothesis is that interviewers restrict respondent behaviors over the field period. Finally, interviewers may not change the prevalence of individual behaviors, but become more efficient in them. This paper examines these three hypotheses using the Work and Leisure Today Survey (WLT; AAPOR RR1=4.7%, n=450). We have paradata containing response time and behavior codes for each of these surveys, including indicators of question misreading, probing, clarification, feedback, adequate answers, and requests for help. Previous research with the Work and Leisure Today Survey has shown that interviewers systematically get faster over the course of the field period. We use cross-classified multilevel models to account for clustering due to interviewers, respondents, and questions, with interviewer and respondent behaviors as the outcomes. We also account for characteristics of respondents, interviewers and questions. Preliminary analyses suggest that indicators of interviewer efficiency such as stuttering while asking the question or having other disfluencies significantly (p.05). This lends support to the third hypothesis. This efficiency finding suggests that increasing study-specific interviewer training on question reading could assist with the "speeding up" effect. We conclude with implications for survey practice and other interviewer training methods.

The Effect of Question Characteristics on Respondent and Interviewer Behaviors

Allyson L. Holbrook, *University of Illinois-Chicago*

Timothy P. Johnson, *University of Illinois-Chicago*

Young Ik Cho, *University of Wisconsin-Milwaukee*

Sharon Shavitt, *University of Illinois at Urbana-Champaign*

Noel Chavez, *University of Illinois at Chicago*

Saul Weiner, *University of Illinois at Chicago*

Respondent behaviors that may indicate difficulties with question answering (e.g., giving a response that does not fit the criteria of the question or asking the interviewer to repeat the question) are hypothesized to be influenced by both question characteristics (e.g., the difficulty of the question or the response format in which they are asked to give their response) and interviewer behaviors (e.g., question reading errors or failure to probe an inappropriate response). In this paper, we use hierarchical linear modeling to analyze behavior coding data to assess the impact of both question characteristics and interviewer behaviors on respondent behaviors to a broad range of health-related questions. We also test the extent to which the effects of question characteristics on respondent behaviors can be explained by interviewer behavior. The data come from a large survey in which Chicago area residents from four racial and ethnic groups (non-Hispanic White, non-Hispanic African-American, Mexican-American,

AAPOR 70th Annual Conference

and Korean-American) were interviewed in-person in the laboratory (in addition to completing several PAPI questionnaires). The in-person interviews involved approximately 150 health-related questions designed to vary across a number of dimensions (e.g., response format, area of health examined, and proxy or first person report). The interviews were audio and video recorded. These recordings were used to code both respondent and interviewer behaviors.

Recording What the Respondent Says: Does Question Format Matter?

Jolene D. Smyth, *University of Nebraska-Lincoln*

Kristen Olson, *University of Nebraska-Lincoln*

Recording respondent answers is a key part of the interviewer's job. Although recording closed-ended questions is a relatively simple task, recording open-ended items requires more skill and working memory capacity for the interviewer. To our knowledge, whether the accuracy of recording for open-ended answers varies if the response format for the interviewer is an open-ended text box compared to a list of items and the interviewer checks all that apply has not been examined. Additionally, how interviewers interact with respondents and how this varies for these two question formats has not been examined. This paper examines the accuracy with which interviewers record respondent answers, with a particular focus on open-ended questions. We use data from the Work and Leisure Today Study (AAPOR RR1=4.7%, n=450), a survey with five open-ended text and 17 open-ended numeric questions. The survey was audio recorded and transcribed. We compare the responses recorded by the interviewer in the survey data to the responses given by the respondent during the interview. We also examine interviewer behaviors such as probing, task-related feedback, and clarification. Preliminary analyses indicate that open-ended text questions have 2 to 3 more conversational turns than any other type of question, that about 9% of interviewer feedback on open-ended questions is task-related, compared to less than 3% on any other type of question, and that about 8% of conversational feedback turns by respondents are digressions, compared to about 6% on all other types of questions. We conclude with implications for interviewer monitoring and training.

Experimenting and Developing Mobile Device Questionnaires

Response Option Order Effects for Different Scale Lengths in Online Surveys Using Different Response Devices

Kristen Martinsson, *University of Gothenburg*

The increased use of online surveys and the increasingly mixed technical devices respondents are using to answer online surveys is the cause of some concern among survey researchers. Despite many efforts to adjust online surveys to smartphones our knowledge is still limited when it comes to how different response scales work on different devices. This paper reports on a large-scale study where not only response option order is varied randomly, but also scale length, the inclusion of don't-know-options, and horizontal vs vertical layouts. All in all, this yields 20 different treatment groups. This was possible due to the unusually large sample size of approximately 40 000 respondents. The main concern is that respondents who use a specific device, e.g. a tablet or a smartphone, might not treat the response scales in the same way. This paper analyzes response option order effects and the large sample size enables us to analyze to what extent these effects are conditional on scale length, layout direction (horizontal vs vertical) and what type of response device respondents are using. The results indicate that horizontal scales do not suffer from response option order effects (primacy effects) to the same

AAPOR 70th Annual Conference

extent as vertical layouts, except for smartphone users. Further, substantial response option order effects appear for smartphone users when scales are longer than 5 scale points.

What They Can't See Can Hurt You: Improving Grids for Online Surveys on Mobile Devices.

Johan K. Thomas, *GfK Custom Research*

Frances M. Barlas, *GfK Custom Research*

Patricia Graham, *GfK Custom Research*

Thomas Subias, *GfK Custom Research*

Measurement in online surveys is often limited by screen real estate. Grids have evolved to efficiently measure multiple concepts using the same response format so that it is completed more quickly than separate presentation. However, grids create significant problems for mobile devices. Even when responsive design is used, not all response options might be visible in a grid on a mobile device without scrolling or changing screen orientation, potentially causing a response bias - participants tend to select what they see rather than what is meant to be seen. With increasing use of mobile devices to complete online surveys, it is critical to reconsider the presentation of grids. We explored two possibilities to improve grids for mobile presentation – reducing the number of response categories and redesigns that reduce the total real estate required. We conducted a study using GfK's KnowledgePanel®, the nation's largest online probability panel. Participants were randomly selected from those who indicated in a prior survey that they used a desktop/laptop and a smartphone or a tablet and randomly assigned to complete the survey on a specified device. We had 3,600 completes complete a 20 minute survey, with 1,200 completing on each device type (desktop/laptop, tablet, smartphone). Some questions had grid presentational formats that varied and were randomly assigned, allowing a comparison of the different grids across devices. We present the effects of these alternate formats on point estimates, overall variance, differentiation, data missingness, response order effects, and response bias by device type. We found that grids can be improved both by reducing the number of response categories and reducing the number and length of elements to evaluate in the visible screen. We provide evidence-based recommendations for improving the presentation and measurement of survey questions that use the same scale and lend themselves to the grid format.

The Impact of Screen Size on Data Quality

Randall Williams, *Westat*

Aaron Maitland, *Westat*

Andrew Mercer, *Westat*

Roger Tourangeau, *Westat*

The number of different mobile and smartphone type devices has grown at an astonishing clip over the last few years and appears to be accelerating. Further, screen sizes for even the most popular devices can vary greatly from as large as 12 inches (Microsoft Surface) to as small as 3.2 inches (Apple iPhone 4S). While there does appear to be a trend toward larger screen sizes within the smartphone market, larger screens may not necessarily be favored by everyone. The issue this raises for survey methodologist is the potential impact on data quality that variations in screen sizes can have. This issue is compounded by the dearth of research on whether mobile devices, such as, smartphones and tablets, affect how respondents answer survey questions. Past research on web questionnaires has demonstrated that differences in placement, spacing, or location of response options affect respondent answers (Tourangeau,

AAPOR 70th Annual Conference

Couper, and Conrad, 2004). Research by Peytchev and Hill (2010) on early mobile devices demonstrated that visual elements outside the visible screen area are likely to be missed by respondents. In this paper, we present the results of an experiment testing the use of an in-person self-administered questionnaire across three randomly assigned interfaces: an iPhone, an iPad and a laptop PC. We show to what extent screen size (visibility) affects how respondents answer questions. We also present result across different question presentation formats, such as grids or standalone questions. Implications for surveys where respondents can choose to respond over mobile devices will also be discussed.

Examining the Impact of Mobile First and Responsive Web Design on Desktop and Mobile Respondents

Douglas Tharp, *Indiana University Center for Survey Research*

Mobile First and Responsive Web Design are two approaches that survey researchers can utilize to improve the experiences of smart phone users, who make up a growing proportion of web survey respondents. The Mobile First approach has a layout optimized for smart phones that also serves as the basis for the desktop design, while Responsive Web Design allows for more dynamic adjustments based on browser size. We conducted experiments with each design within the past year. Both experiments were attempts to better meet the needs and expectations of mobile respondents with designs that could also be employed for desktop respondents with minimal differences in layout. For each, a random sample of respondents was assigned to receive the experimental version of each survey, while other respondents were assigned to a more traditional, desktop-focused design. The Mobile First experiment was conducted first. Smart phone users who were assigned the Mobile First layout were less likely to break off, and had lower duration times than smart phone users assigned to the traditional design. They also gave the Mobile First design higher ratings in a short post-survey evaluation questionnaire. Desktop users, however, had longer duration times when completing the Mobile First design when compared to the traditional design. They also rated the Mobile First design lower on professional appearance, and commented on excessive need for scrolling. We found no significant differences in response distribution among the layout versions and devices. Our second experiment was designed to address these specific concerns of desktop while maintaining the positive benefits for smart phone users. When this additional experiment is completed, we will compare both designs, considering data quality, response rates, evaluation scores, cost, and development time.

What Is The Impact of Smartphone Optimization on Long Surveys?

Kevin Brooks, *Center for Postsecondary Research, Indiana University*

Shimon Sarraf, *Center for Postsecondary Research, Indiana University*

James S. Cole, *Center for Postsecondary Research, Indiana University*

Each year, an increasing number of college student survey respondents are accessing online surveys using smartphones, instead of desktop computers (Sarraf, Brooks, & Cole, 2014). In 2011, about 4% of respondents to the National Survey of Student Engagement (NSSE) used a smartphone, but by 2014 the proportion increased to about 18%. The widespread adoption of smartphones among college students has prompted discussions among some institutional and higher education researchers about data quality and appropriate survey formats. Some research suggests that there is little or no substantive difference for users accessing web surveys on tablets as compared to desktop machines, but the user experience on a smartphone is recognized to be significantly different than the other devices (Buskirk & Andrus, 2012). This has caused some researchers to optimize their online surveys for smartphones, however

AAPOR 70th Annual Conference

research is limited about how this may affect survey data quality. Using results from a five-institution experiment using the 2014 National Survey of Student Engagement (NSSE) the current study details the impact that smartphone optimization has on a survey with over one-hundred questions. Study research questions center on how various data quality indicators are effected by optimizing a survey for smartphones, including item skipping, completion time, scale factor structure, and response option differentiation (straight-lining). Based on their recent experience, presenters will offer insights into developing a smartphone-optimized version of a relatively long survey.

Mobile Devices for the Collection of Sensitive Information

Jennifer Tourangeau, *Westat*

Doug Williams, *Westat*

Aaron Maitland, *Westat*

Andrew Mercer, *Westat*

This paper examines mobile devices—both smartphones and tablet computers—as vehicles for survey data collection. The appeal of these devices for survey researchers is obvious. Because they are lightweight and relatively inexpensive, they make it easier to collect data using such existing survey modes as computer-assisted personal interviewing (CAPI). The use of smartphones and tablets in surveys does raise several issues, including the perceived privacy of data collected on these devices. Respondents are willing to reveal sensitive information about themselves when a computer administers the questions. But it is unclear whether respondents will display the same level of candor when the survey is administered over the Internet on a tablet computer or a smartphone. We examined this issue in a realistic field experiment. The experiment compares the same three data collection platforms (smartphones, tablets, and laptops) as vehicles for collecting sensitive information, using items on alcohol use, smoking, and illicit drug use drawn from national surveys, such as the National Survey of Drug Use and Health. The data collection device is crossed with self- versus interviewer administration in a 3 x 2 factorial experiment. The key outcome in this experiment is the level of reporting potentially embarrassing information. The experiment also asks respondents to assess the sensitivity of the questions, which can vary as a function of the method of data collection. Finally, the experiment will compare rates of missing data, response rates, and breakoffs as a function of the device and method of administration.

Sunday, May 17
8:00 a.m. – 9:30 a.m.
Concurrent Session J

Mini-Conference Panel: Using Latent Class Analysis to Evaluate Survey Quality

Mini-Conference Panel: Using Latent Class Analysis to Evaluate Survey Quality

Paul Biemer, *RTI International*

Latent class (or structure) analysis (LCA), a theory for detecting unobserved variables, was developed by Paul Lazarsfeld (1950). According to Lazarsfeld, an unobserved variable could be constructed by taking into account the interrelationships among observed or “manifest” variables assumed to be related to the unobserved variable. The mathematics underlying this theory was extended by Lazarsfeld and Henry (1968) and Goodman (1974). Over the last twenty years, survey methodologists have used LCA to study measurement error in surveys. Often LCA is used to evaluate the error properties of survey questions by examining the interrelationships between items purportedly measuring the same construct, but it also has been used to evaluate survey design or survey measurement more broadly. For example, LCA has been used to estimate the respondent’s level of effort using paradata. This session presents the results from several methodological studies using LCA. These studies not only illustrate the different ways LCA can be used to evaluate survey quality but also the differences in opinions on the utility of LCA in this area.

Latent Class Analysis: What, Why and How?

Paul Biemer, *RTI International*

Latent class analysis refers to a statistical framework for modeling categorical data along with their classification errors. The general latent class model postulates that the true value of a variable is unobservable (latent) while a survey response may constitute a single indicator of this latent variable. The LC model parameters include the target population proportions for a categorical variable to be estimated in the survey and the probabilities of misclassification probabilities (for e.g., false positive and false negative, for dichotomous response variables) for measuring the variable. Survey item reliability and construct validity as well as estimator bias can be defined and interpreted in this context. One advantage of viewing survey classification error model as a LCM is the availability of general software for estimating the model parameters including the error components. However, the assumptions of the traditional LCM can be somewhat restrictive. An even more general model can be obtained by viewing the LCM as a type of log linear model with latent variables. In doing so, a wide range of error structures and error evaluation designs can easily be discussed and analyzed using log-linear modeling notation and methods. This presentation will provide a general introduction to latent class analysis describing what it is, the types of research questions for which it is ideally suited to answer and some cautions to regard in its applications. LCA for both cross-sectional and panel survey analysis will be discussed. Methods for analyzing complex surveys designs with missing data will also be covered. Several illustrative applications of LCA to survey data will be briefly described.

Survey Measurement Errors That Are Not All One-Way: Applying the Latent Class MTMM Model

Daniel Oberski, *Tilburg University*

Perhaps one day we will figure out how to ask perfect survey questions. In the meantime, survey analyses are biased by random and correlated measurement errors, and evaluating the extent of such errors is therefore essential, both to remove the bias and to improve our question design. When there is no gold standard, these errors are often estimated using multitrait-multimethod (MTMM) experiments or longitudinal data by applying linear or ordinal factor models, which assume that (latent) measurement is linear and that the only type of method bias is one that pushes the answers monotonely in a particular direction—that of acquiescence, for example. However, not all measurement is linear and not all method bias is monotone. Extreme response tendencies, for example, are nonmonotone, as are primacy and recency effects, which act on just one category. Just as the monotone kind, such method effects will also lead to spurious dependencies among different survey questions, distorting their true relationships. Diagnosing, preventing, or correcting for such distortions therefore calls for a model that can account for them. For this purpose I will discuss the latent class MTMM model (Oberski 2011). In it, a latent loglinear modeling approach is combined with the MTMM design to yield a model that provides detailed information about the measurement quality of survey questions while also dealing with nonmonotone method biases. I will discuss the method's assumptions and demonstrate it on a few often-used survey questions. Standard software for latent class analysis can be used to estimate this model, so that evaluating the extent of nonlinear random and correlated measurement errors is now a reasonably user-friendly experience for survey researchers.

Using a Two-Part Markov Latent Class Model to Examine the Quality of Consumer Expenditure Reports

Brian Meekins, *Bureau of Labor Statistics*

Clyde Tucker, *American Institutes for Research; CNN*

The Consumer Expenditure Survey is a rotating panel survey where households are interviewed in four consecutive quarters about their purchases and expenditure for the prior three months. Given the comprehensive nature of the survey, where a household is asked to provide detailed expenditure for a large number of different types of commodities, the survey is thought to be burdensome and reports of expenditure measured with significant error. Previous research on the Consumer Expenditure Survey has applied both Latent Class Models that utilize survey paradata to develop classes of reporters based on quality (Meekins, Beimer, Tucker 2011), and Markov Latent Class Models (Hidden Markov Models) to identify unreported purchases (Tucker, Meekins, Beimer 2011). Although showing great promise, the Markov Latent Class Models (MLCA) focuses exclusively on the pattern of reporting a purchase while not examining the amount of the reported expenditure. While, much of the error in reporting expenditure most likely occurs from a household forgetting a purchase, there is likely some degree of error that occurs from misreporting the expenditure amount. The current research describes and estimates a two part model similar to Olsen and Schafer (2001) with an underlying Markov Process. In this model the latent probability of an actual purchase is estimated at each of the four quarters, using the reported purchase as an indicator. The mover-stayer specification is then added that describes the entire purchase behavior where a household can be classified under a latent variable as either a stayer purchaser (a purchase in every quarter), stayer non-purchaser (no purchase in any quarter), or mover (all others). Under the Mover-stayer specification latent expenditure (using reported expenditure as an indicator) will be estimated for

AAPOR 70th Annual Conference

those classified as mover or stayer purchasers. In this way the model will address error from reported purchases and error in expenditure reports.

Using Latent Class Models to Identify Problem Items: An Empirical Appraisal

Roger Tourangeau, *Westat*

Frauke Kreuter, *University of Maryland, JPSM*

Ting Yang, *University of Michigan*

This talk discusses three attempts to use latent class models to assess the measurement properties of survey items. In the first study, latent class models were fit to items measuring academic problems experienced by a sample of alumni when they were undergraduates. The survey answers were also compared to the respondents' transcripts, providing true values for assessing the accuracy of their responses. The latent class estimates of the error rates for the items showed good qualitative agreement with true value estimates, but poor quantitative agreement. For example, the latent class analysis (LCA) identified the worst item in terms of error rates, but the LCA estimates of the error rates were not close to those from the true value estimates. In the second study, various indicators of the quality of a set of items were available, including cognitive testing results, reliability measures, and LCA estimates of error rates. In addition, some of the items were written to include "planted" problems. There was poor agreement across the indicators as to which items performed poorly and the items with planted problems were not always identified as problematic. In a final study, very similar questions were asked in a national survey of women first under CAPI and later under ACASI. LCA models give implausible estimates of the error rates under the two modes. Overall, error rate estimates from LCA appear to be useful, but only suggestive.

Comparison of Ratings and Rankings for Measuring Work Values Preferences: A Latent Class Segmentation Approach

Guy Moors, *Tilburg University*

Ingrid Vriens, *Tilburg University*

John Gelissen, *Tilburg University*

A continuing discussion in sociological survey research concerns whether social values should be measured using either a rating or rather a ranking response format. The form-resistant hypothesis states that differences in the latent preference structure revealed by both approaches should be small when typical features of each format are considered. Previous research, however, has shown mixed results. We suggest that adopting a latent class segmentation approach helps to explain these mixed results: It may identify segments in the population with a similar item preference structure – regardless of whether rankings or ratings are used –, as well as segments that are linked to one format only. We apply our approach to a Dutch nationally representative survey on work values with a split-ballot design. In both the rating and ranking assignment we find two segments reflecting the intrinsic and extrinsic work values preference structure. At the same time other preference structures defined segments that differed between modes. In line with the form-resistant hypothesis the results suggest the same latent preference structure has guided particular segments in a population to respond similarly to rating and ranking questions.

Partisan Identity and Ideological Formulation

Ideological Self-Identification, Political Values and Partisanship

Jocelyn Kiley, *Pew Research Center*

Scott Keeter, *Pew Research Center*

Alec Tyson, *Pew Research Center*

Measures of ideology in public opinion research often rely on self-identification, typically asking respondents whether they are liberal, conservative or moderate. But some evidence suggests that these terms, while commonly used by people who study American politics, may be less familiar to the general public. How does ideological self-identification correspond to ideological placement using measures based on political values? And what is the relationship between different measures of ideology and partisanship? Overall, Americans are more likely to self-identify as conservative (36%) than liberal (23%); yet on a scale of political values more Americans are mostly or consistently liberal (34%) than mostly or consistently conservative (27%). Using a Pew Research Center survey of over 10,000 respondents, along with a subset of these respondents who were recruited to take additional surveys over the course of the year, we further explore the relationship between ideological self-placement and ideology as derived from measures of underlying political values. Respondents were asked ideological self-identification two different ways at different time points: a standard five item scale in a telephone survey, and a visual scale with more response options on a web/mail survey. In addition, respondents were asked a full battery of political values that allowed for the development of a scale of ideological consistency. This paper will investigate the relationship between these different measures of ideology, across demographic and political groups. Additionally, it will explore how these measures correlate with measures of political knowledge, political interest and engagement, partisanship, and views of the political parties (including views about the ideological placement of the parties).

The Gendered Roots of Political Engagement

Laura Lazarus Frankel, *Duke University*

This paper explores how a person's gender (masculine and feminine) traits affect their level of political ambition and likelihood to run for public office. Using original survey data from the 2012 Cooperative Congressional Election Study (CCES), I run a factor analysis to determine each respondent's unique level of masculinity and femininity. I then use these "gender scores" to predict political confidence and a variety of attitudes and behaviors required of political candidates. While many of the results are consistent with expectations, several run counter to existing theories and popular wisdom about how gender affects political ambition. This research expands our understanding of what propels men and women to engage with politics and seek positions of political leadership.

Political Polarization and Public Attitudes About Science: Beyond Climate Change

Cary Funk, *Pew Research Center*

Jocelyn Kiley, *Pew Research Center*

Public opinion related to science topics has become increasingly politicized over the past decade, particularly when it comes to climate change. For example, the 2014 Exit Polls showed voters' views about climate change were strongly linked to House Vote; the political divide over

AAPOR 70th Annual Conference

climate change rivals that of any issue considered on the exit polls. Beyond a handful of hot button issues, however, few have addressed the extent to which political differences underlie public attitudes about science issues, writ large. We address this gap with a new representative telephone survey with roughly 2,000 U.S. adults from the Pew Research Center that includes attitudes about a wide range of science topics. We examine the extent to which political differences explain public attitudes about the contribution of science to society, support for government research funding, views about regulation, and opinions across a range of science topics including: space exploration, GMOs, bioengineering of human organs, genetic modifications, animal research, nuclear power, fracking, and offshore oil drilling. We use multivariate analyses to test for the independent effect of political party, while controlling for factors such as religious affiliation, science knowledge and training, gender, age, socioeconomic background, race and ethnicity. We show that while there are a number of issues, such as climate change, where political groups hold divergent views, on other topics political differences are modest. We also find that both Republicans and Democrats tend to see science as contributing to society but that differences emerge in their views about government research funding and regulation, in keeping with broader ideological differences about the role of government. These findings challenge the notion of a broad-based politicization of science and lay out a deeper understanding of the role of politics in public attitudes about science.

Why Americans Deny their Partisan Identities and What It Means for Politics

Samara Klar, *University of Arizona*

Yanna Krupnikov, *Stony Brook University*

Each election year, media describe the large segment of the electorate who identify as independent— nearly forty percent of Americans at last count – as pivotal for electoral outcomes. Academics, on the other hand, conclude that independents are merely “undercover partisans” – people who secretly hold clear partisan beliefs and consistent preferences for partisan candidates. Yet we are left with a nagging question: If independents are simply “undercover partisans,” why are they so motivated to intentionally misrepresent their own partisanship? Even more importantly, could the motivations that lead individuals to avoid partisan labels also lead them to alter their political behaviors? We attribute the growth in independents to two colliding factors: a growing social stigma against partisans of all stripes and an individual-level concern for presenting oneself in a positive light. Using a large panel survey dataset across two election years, we demonstrate that people who are more concerned about the impressions they make on others are significantly more likely to tell survey researchers that they identify as independents. But the consequences extend far deeper than mere survey responses. By tracking respondents’ political action, we also show that the motivations leading people to identify as independent on surveys also discourage them from engaging in any act that might betray their partisan identity: actions that include displaying partisan pins or stickers and discussing political preferences with others. Fed a steady diet of media coverage of partisan disagreement, Americans have grown embarrassed of their own partisan attachments. In turn, they deny to pollsters, party activists, friends, and even themselves their true partisan tendencies, instead choosing to go “undercover” as independents.

Sampling Local Areas With Cell Phones

Use of Small Area Analysis in Survey Analysis for Health Policy: Example from the 2015 Ohio Medicaid Assessment Survey

Daniel Joseph Weston, *The Ohio Colleges of Medicine Government Resource Center*

Marcus Berzofsky, *RTI International*

bo Lu, *The Ohio State University*

Timothy Sahr, *The Ohio Colleges of Medicine Government Resource Center*

Lance Couzens, *RTI International*

Chengzhou Zhang, *The Ohio Colleges of Medicine Government Resource Center*

The Ohio Medicaid Assessment Survey (OMAS) provides health policy and health assessment information to Ohio's public and private health insurers and providers. A major emphasis for this survey is assessment of Medicaid and Medicaid-eligible populations – specifically profiling these populations in relation to activities and benefit offerings associated with the Affordable Care Act (ACA). The 2015 OMAS is a mixed collection mode (landline and cell phone) probability sample of 34,000 adults, with 8,800 adults providing proxy responses for a child in the household. The OMAS sample provides a robust sampling base for counties of approximately 50,000 population or greater but has difficulty supporting reliable direct survey estimates for small rural counties. To enable county-level estimations of key policy variables including the uninsured and underinsured adults, Medicaid expansion population, health insurance exchange population, and uninsured children, researchers will employ small area estimation (SAE) to obtain county-level estimates of ACA impact. Using Ohio's 88 counties as a base unit of analysis, this paper will compare estimates yielded from SAE to direct estimates of counties with adequate sample. Discussion will center upon the uses of SAE for ACA-associated state-specific analyses. Due to a plethora of auxiliary information at the area or county levels, it is natural to follow Rao's (2003) Basic Area Level (Type A) Model for SAE.

Reducing Cell Phone Coverage Bias in Geographically Targeted RDD Samples By Weighting for Residential Mobility

Tara Merry, *Abt SRBI*

Stephen Immerwahr, *New York City Department of Health and Mental Hygiene*

Michael Sanderson, *New York City Department of Health and Mental Hygiene*

Andy Weiss, *Abt SRBI*

Michael Battaglia, *Battaglia Consulting Group, LLC*

Rachel Martonik, *Abt SRBI*

Josh Appelbaum, *Abt SRBI*

The geographic portability of cell phones poses a challenge for RDD surveys below the national level. As the cell phone only (CPO) population grows, and many CPO users maintain phone numbers with no geographic link to their current home, there is an increasing potential for "in-area, out-of-frame" coverage error. We examine the impact of weighting for how recently someone has moved and explore whether weighting recent CPO movers, in particular, has potential to reduce bias due to the exclusion of CPO users who have non-local cell phone numbers. The 2014 New York City (NYC) Community Health Survey (CHS) included two questions about mobility based on questions in the American Community Survey (ACS). Compared to ACS estimates for NYC, recent in-mover (past 12 months) NYC residents appear to be under-represented in the CHS sample, to differ significantly from other NYC residents, and are much more likely to be CPO. Therefore, if residential mobility (non-mover, moved within

AAPOR 70th Annual Conference

NYC, and moved from outside NYC) is added as a raking margin to the weighting, it will increase the weights of CPO adults who were living outside of NYC 12 months prior to the interview. If recent CPO movers in the sample are similar to the “in area, out of frame” adults, then residential mobility may be a reasonable proxy to help correct for frame non-coverage in small area RDD telephone surveys. The effect of adding residential mobility to the raking for the 2014 CHS will be analyzed to understand whether this may reduce bias caused by the “in area, out of frame” problem.

Methods to Account for Classification Error in County Assignment Based on Rate Center in a Periodic Survey

Caroline W. Blanton, *RTI International*

Marcus Berzofsky, *RTI International*

Bo Lu, *Ohio State University*

Kimberly Peterson, *RTI International*

Lance Couzens, *RTI International*

Jamie Ridenhour, *RTI International*

Timothy Sahr, *Ohio State University*

Robert Ashmead, *Ohio State University*

Amy Ferketich, *Ohio State University*

Thomas Duffy, *RTI International*

State or local-area based surveys are often designed to make estimates at the county or county-group levels. Under a traditional random digit dial (RDD) design, the telephone exchange of a landline number could be used to accurately identify the county for which the associated household resides. However, initially, no good analogous data methods existed for the cellphone frame. This required survey methodologists to only draw random samples of cell numbers from the entire state; thus, making it impossible to target areas within a state. To overcome this shortcoming, sample vendors, such as Marketing System Group (M-S-G), have recently been able to identify a cellphone number's rate center and determine the county from which the rate center most likely resides. Speizer, et. al. (2013) showed that rate centers can be used to identify a respondent's county of residence, but may have classification error rates as high as 30%. These high classification error rates make it difficult to accurately devise a sample allocation using a cellphone frame. However, for periodic surveys, such as the Ohio Medicaid Assessment Survey (OMAS), past iterations of the survey can be used to estimate not only the overall classification error rate, but the probability that the cellphone owner will actually reside in the county assigned. This paper presents the success to which the 2015 OMAS utilized this information to allocate its sample. Building on the methods proposed in Speizer, et al. (2013), we will demonstrate how survey designers can use past survey information to inform a new survey's cellphone sampling strategy and the extent to which we were able to accurately allocate a cellphone sample to target respondents at the county and county-group level.

Cellphone Sampling at the State Level; Geographic Accuracy and Coverage Concerns

Stephanie Kafka, *The Gallup Organization*

Brad Hoffmann, *The Gallup Organization*

Manas Chattopadhyay, *The Gallup Organization*

The prevalence of cellphone users amongst the U.S. population has increased steadily, necessitating a higher percentage of cellphone only and dual user (cellphone and landline

AAPOR 70th Annual Conference

households) in survey samples to provide appropriate coverage of the U.S. population. The increased percentage of cellphone numbers as a proportion of the total telephone sample frame has complicated survey sample design, especially at the state level, because the phone service could have been acquired in a different state than the person currently lives. Thus some respondents that have moved into the state will never be included when selecting a state-level sample using traditional sampling techniques. Researchers have established that approximately 12% of cell-phone only adult households reside in a different state than the state associated with their cellphone number (Christian et. al. 2009). In their research Skalland et. al. (2012) demonstrated how this inaccuracy varies tremendously by state, however a deep evaluation of all cellphone households including cellphone only and dual users has not been conducted. Gallup seeks to build upon this growing body of research utilizing more recent data collected as part of the 2013 State of the States Study of more than 30,000 adult households. In this national study Gallup categorized respondents based on their self-reported state, however the state appended to the sample allows researchers to evaluate the geographic accuracy of the cellphone service for all sampled households, including cellphone only and dual user households. Gallup will provide important information about the demographic distribution of cellphone only and dual-user households that have moved from the state associated with their cellphone service, to establish how excluding these respondents from the sample frame can affect the final distribution of the data. Finally, Gallup will establish which states, if any, can benefit from sampling adjacent states to provide better coverage of their current population.

Polling Potpourri: Tobacco, Alcohol and Firearms

Views About Marijuana – Analyzing the Correlates of Support for Legalization

Alec Tyson, *Pew Research Center*

Jocelyn Kiley, *Pew Research Center*

One of the most dramatic shifts in public opinion over the past decade has been the rise in support for the legalization of marijuana: In 2000, just 31% of Americans favored marijuana legalization; now more than 50% do so. Some of the dynamics of this shift in opinion have been widely reported, including generational differences in views of marijuana legalization. But growing support for legalization spans most generations. This analysis will leverage two 2014 Pew Research Center national RDD surveys – a 10,013 case Polarization survey and the 1,821 case February political survey – to examine the correlates of support for legal marijuana in greater depth. The paper will look at how a range of attitudes on marijuana – from its role as a gateway drug, its impact on personal health, and attitudes about its use in public spaces – are tied to overall views of legalization. In addition, it will leverage the breadth of concepts measured on our Polarization survey, and over several subsequent interviews with the same respondents, to explore how views of legal marijuana are related to political partisanship, ideology and other policy issues, as well as demographic and personality characteristics.

Gun Ownership in the United States: Measurement Issues and Trends

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

Jaesok Son, *NORC*

Gun violence is a serious problem in the United States. In 2011, 478,400 violent crimes were committed with a firearm. While down dramatically since the 1990s, the rate of non-fatal firearm victimizations rose from 2008 to 2011. Firearms deaths from all causes (homicides, suicides,

AAPOR 70th Annual Conference

accidental, and undetermined) averaged over 31,000 annually in 2005-2011. Non-fatal gunshot injuries totaled 81,396 in 2012; a rise in the injury rate per 100,000 from 20.5 in 2002 to 25.9 in 2012. Given the magnitude and seriousness of gun violence, it is important to have accurate and reliable information on the possession and use of firearms in the United States. This report examines one crucial element, the level of and trends in household and personal gun ownership. First, the report considers methodological issues concerning the measurement of gun ownership. Second, it examines trends in gun ownership. Third, it evaluates the nexus of these two factors, the impact of methodological issues on the measurement of trends gun ownership. Finally, it considers what ancillary trend data on crime, hunting, household size, and number of guns available suggest about trends in gun ownership. Two primary sources were analyzed. First, the General Social Surveys (GSSs) conducted by NORC at the University of Chicago. Second, drawing mostly on the IPOLL archive of the Roper Center for Public Opinion Research, a database of 415 United States, national polls that asked about gun ownership was compiled. This covers questions from 1959 to early 2013. It includes 364 questions about household gun ownership and 51 questions about personal gun ownership. This contains data over a dozen other data collectors. In addition, other sources of information from the research literature on gun ownership and from relevant data on crime, hunting, household size, number of guns available, and other related matters have been utilized.

When Do Religion and Science Conflict?

Besheer Mohamed, *Pew Research*

Cary Funk, *Pew Research Center*

Becka Alper, *Pew Research Center*

Is there an inherent conflict between science and religious belief? There are well-documented opinion differences among religious groups on a handful of science topics including beliefs about evolution and stem cell research. Little is known about the extent to which religious group differences underlie views about other science topics. We use a new representative survey by the Pew Research Center to examine the extent to which religious group differences explain public attitudes about science across a wide range of science topics including: views about evolution, bioengineering of human organs, genetic modifications, and opinions about animal research, climate change, and energy issues. We use multivariate analyses to test for the independent effect of religious group, while controlling for factors such as political orientations, science knowledge and training, gender, age, socioeconomic background, race and ethnicity. The survey also addresses the public's perceptions of conflict between religion and science. Findings show that those who are less religiously committed tend to perceive more conflict between religion and science than do those higher levels of religious commitment. Opinion differences among religious groups while typically quite large when it comes to topics with a clear link to views about the origins of life tend to be modest or not statistically significant on a range of other science topics. Thus, the findings show a surprising degree of consensus among religious groups about many science topics while at the same time underscoring areas where there are fundamental disagreements among religious groups in U.S. society today.

AAPOR 70th Annual Conference

A Population-Based Smartphone Survey on Tobacco Use

Sean Hu, *Centers for Disease Control and Prevention*

Naomi Naomi Freedner-Maguire, *ICF International*

James Dayton, *ICF International*

Linda Neff, *Centers for Disease Control and Prevention*

Smartphone technology can help engage respondents in public health surveillance by providing web access that is rapid, convenient, and confidential. The purpose of this feasibility study was to assess the use of smartphones as a mode for administering a population-based web survey about tobacco use behaviors. Starting with a national sample of 25,000 random-digit-dial cell phone numbers, 1446 respondents aged 18 to 65 years successfully completed a screening interview for eligibility to participate in the study. Of the screened sample, 1068 respondents were deemed eligible as smartphone users, of whom 528 (49%) agreed to participate in the study. During June, 2013, study participants were texted links to complete two web surveys with 18 questions each over the course of two weeks. The contact rate for both web surveys was approximately 80% and the response rate for both was 28%. Smartphone survey data were compared to data from the 2012-2013 National Adult Tobacco Survey (NATS), a random-digit-dial landline and cell phone survey of U.S. adults. The study assessed: 1) demographic characteristics of respondents by survey modes using pairwise contingency tables; 2) bivariate relationships between survey modes and tobacco use behaviors; and 3) multivariate logistic regression models for each type of tobacco use, adjusting for age and race/ethnicity, to examine whether survey mode affects responses[A1]. Smartphone respondents reported significantly lower rates of using cigarettes, cigars, e-cigarettes and water pipes than those in the NATS cell phone survey; however, the only difference observed between the smartphone and NATS landline surveys was reported use of cigars. The study findings provide insights into the feasibility of using smartphones for public health surveillance. The findings reveal that a population-based, tobacco-related interviews can be conducted by smartphone; however, additional research is critical to ascertain the validity and reliability of the data that are obtained.

The Powerful Effects of Social Resources on Community Resilience: An In-Depth Study of Twelve Neighborhoods' Recovery from Superstorm Sandy

Daniel Malato, *The Associated Press-NORC Center for Public Affairs Research*

Trevor Tompson, *The Associated Press-NORC Center for Public Affairs Research*

Jennifer Benz, *The Associated Press-NORC Center for Public Affairs Research*

Becky Reimer, *The Associated Press-NORC Center for Public Affairs Research*

David Sterrett, *The Associated Press-NORC Center for Public Affairs Research*

Emily Alvarez, *The Associated Press-NORC Center for Public Affairs Research*

With the frequency and severity of natural disasters on the rise, understanding what factors lead some communities to be more resilient than others is critically important for society. Superstorm Sandy heavily affected large areas of New York and New Jersey in Oct. 2012. Two years later, the recovery from the storm is not universal, and this raises an important research question: What leads some communities to recover more quickly and easily from natural disasters than other communities? To address this question, the AP-NORC Center, with funding from the Rockefeller Foundation, extended previous research into disaster recovery with an in-depth study looking at how neighborhood social resources impact community resilience. This 2014 study utilized an address-based sampling approach in 12 neighborhoods in New York and New Jersey heavily impacted by Superstorm Sandy to conduct a multi-mode survey with 1,009 residents by web, phone, and in-person interviews. While many people assume disaster

recovery is based solely on neighborhood wealth or outside financial assistance, this study provides a comprehensive look at the important ways that social factors contribute to community resilience. This analysis demonstrates that neighborhood cohesion, collective efficacy, social exchange, and general trust vary across neighborhoods and are all strongly associated with how residents perceive the resilience of their community, regardless of neighborhood socioeconomic status and the impact of the storm. Social resources can make neighborhoods with fewer financial resources more resilient, and an absence of social resources can make it more difficult for wealthier neighborhoods to recover from disasters. The 2014 study incorporates interviews, focus groups, and neighborhood observations with the survey's findings, and shows that the deficits in financial resources that undoubtedly influence resilience can be overcome in part by the presence of social resources.

Topical Issues in Measuring Sexual Violence and Victimization: Effects of Questions, Interviewers and Context

Interviewer Effects and the Administration of Sensitive Behaviorally Specific Questions

Reanne L.M. Townsend, *Westat*

Aaron Maitland, *Westat*

Antonia Warren, *Westat*

David Cantor, *Westat*

Many surveys of rape and sexual assault (RSA) rely on behaviorally specific questions (BSQ), which use language that is explicit and highly sensitive. BSQ also tend to be complex, addressing multiple concepts such as use of force and non-consent (Fisher & Cullen 2000). Many RSA surveys that use BSQ, including the National Violence Against Women Survey (NVAW) and the National Intimate Partner and Sexual Violence Survey (NISVS), are administered by interviewers over the phone. Despite the sensitive and complex nature of BSQ, there is limited work examining how and to what extent interviewer behavior and interaction with the respondent contributes to survey error for these types of questions. The purpose of this study is to analyze interviewer and respondent behavior during administration of BSQ in a computer-assisted telephone interview on RSA. The data used come from a feasibility study done in connection with the National Study of Health and Safety (NSHS), a Bureau of Justice Statistics study testing alternative measures of RSA. In this paper we will address 3 questions related to interviewer effects and BSQ. First, we investigate whether interviewer and respondent behaviors differ between interviewers. This will use approximately 100 audio-recorded interviews that have been coded for interviewer and respondent behaviors; including whether the question is read correctly, whether the interviewer probes, and whether the respondent requests clarification. The second question is whether these effects differ for BSQ that vary in sensitivity. The analysis will compare BSQ items that ask about respondents' personal experiences with RSA to a hypothetical RSA scenario presented in a vignette. Third, we will investigate whether there is significant variation between interviewers in the likelihood that a respondent reports a personal RSA incident. All of these analyses will rely on both descriptive and multi-level models to characterize and assess interviewer variation.

AAPOR 70th Annual Conference

Comparing Three Measures of Sexual Assault

David Cantor, *Westat*

Darby Steiger, *Westat*

Shannan Catalano, *Bureau of Justice Statistics, U.S. Department of Justice*

Reanne Townsend, *Westat*

Current “best practice” to survey measurement of rape and sexual assault (RSA) is administration of behaviorally specific questions (BHQ) (NRC, 2013). These are characterized by explicit language to detect and classify self-reports as RSA. Prior applications of the BHQ methodology have used a single question to classify and count events as RSA. This is recognized as problematic because of the complexity of these items (Cook, et al., 2011). At last year’s AAPOR meeting, we presented results based on a series of cognitive interviews that BHQ are subject to multiple interpretations and measurement error (Cantor, et al, 2014; Steiger, et al., 2014). The proposed presentation extends the results from the cognitive interviews to a feasibility survey administered to approximately 200 females in the spring of 2014. This research is being conducted by the Bureau of Justice Statistics as part of a redesign of the National Crime Victimization Survey. The analysis first asks whether it is possible to administer follow-up questions after BHQ have been administered, including narratives. There is some concern that follow-ups might be too burdensome or sensitive. Second, how do three different measures of RSA compare: 1) a single BHQ administered as a screener, 2) use of follow-up questions asking details related to what happened and 3) self-reported narratives of what happened during the event. The analysis indicates that women were able and very willing to provide details about specific events and many provided verbatim reports of what happened. Reporting on multiple incidents did pose problems. Comparing the three different methods to describe the event indicates inconsistency in characterizing what happened, especially for incidents involving coercion, attempted RSA and alcohol-related events. There is error associated with each measure and more than one approach is needed to accurately characterize what happened.

Improving Recall of Crime for the National Crime Victimization Survey

Rene Bautista, *NORC at the University of Chicago*

Lisa Lee, *NORC at the University of Chicago*

Pamela Loose, *NORC at the University of Chicago*

Stephanie Poland, *NORC at the University of Chicago*

Shannan Catalano, *U.S. Department of Justice Bureau of Justice Statistics*

NORC at the University of Chicago conducted research on behalf of the Bureau of Justice Statistics (BJS) to examine how adding contextual information to the National Crime Victimization Survey (NCVS) can improve data quality. In this research, we examined the effects of a memory aid, called the Enhanced Contextual Priming (ECP) module, on the recall of crime. The ECP memory aid was designed and tested first in a series of cognitive interviews and then in a larger field test. The memory aid included questions on respondents’ feelings of safety at home and at the places they go and their level of trust in others. We hypothesized that the memory aid would facilitate respondent reporting of crimes and that the broader context could increase respondent engagement, leading to improved response rates. NORC conducted a nationwide telephone field test with 2000 respondents. Respondents were randomly assigned to one of two conditions. The control group received a streamlined version of the traditional NCVS instrument, which included the full crime screener and an abbreviated incident report. The treatment group received the same streamlined NCVS instrument preceded by the ECP

AAPOR 70th Annual Conference

module. The results suggested that the ECP improved recall overall, and specifically for property crimes. Differences in response to the ECP were observed for victims vs. non-victims. Further, respondents adopting an optimizing approach to the survey, as evidenced by their fuller responses to the ECP, reported more crimes than those who satisficed. Although it was expected that contextual priming could increase respondent engagement and improve response rates, no effect on response rates was observed. Overall, the ECP showed promise as a tool for collecting information on correlates of crime from both victims and non-victims and for enhancing recall of crime victimization.

Do the Self-Report Data Reflect the Real Burden of Lifetime Exposure to Sexual Violence Among Girls Aged 13 to 24 Years in Malawi?

Amy Z. Fan, *Centers for Disease Control and Prevention*

Howard Kress, *Centers for Disease Control and Prevention*

James Mercy, *Centers for Disease Control and Prevention*

Objectives: to quantify the different risk of lifetime exposure to sexual violence (SV) reported by girls aged 13 to 24 years; calibration is sought to adjust the self-recall bias. Background: presumably, the lifetime experience of SV among girls and young women should increase with age. However, self-reported data from Violence Against Children Survey (VACS) in Malawi indicated that the prevalence of SV is quite similar between two age groups (13-17 yrs girls, 34.3% and 18-24 yrs young women, 34.1%). Methods/Design: Data from a nationally representative sample of females aged 13 to 24 yrs in Malawi in 2013 was analyzed. The response rate was 83.4%. SV was defined as unwanted touching, attempted, pressured or physically forced sex. Survival analysis was used to model whether age grouping influences the reported age at the first SV incident. The distribution of four types of SV was compared between two age groups. A calibration was performed to obtain an estimate of the lifetime SV prevalence which is closer to the real burden. Results: First incident of SV is reported to occur 3.8 years younger among girls than young women. The risk of experiencing SV increased three times for girls compared to young women (Hazard Ratio=3.12). Among females who have experienced SV, young women were more likely to report more serious SV types (forced or pressured sex, 41.2%) as their initial SV experience than girls (17.8%). The calibrated weighted prevalence of lifetime SV was 56.6% which was much higher than the non-calibrated estimate (33.6%). Conclusion and implications. The findings suggest that young women appeared to “forget” less serious SV which may have occurred much earlier than their alleged initial experience of SV. Prevalence estimates of lifetime SV solely based on self-recall can be seriously under-estimated especially among females older than 18 years.

Designing a Companion Survey to the NCVS Using a Mail Questionnaire

Sherman Edwards, *Westat*

Pamela Giambo, *Westat*

Pamela Broene, *Westat*

Michael Planty, *Bureau of Justice Statistics*

J. Michael Brick, *Westat*

Sharon Lohr, *Westat*

The Bureau of Justice Statistics and Westat have a cooperative agreement to develop a Companion Survey (CS) to the National Crime Victimization Survey (NCVS), to support estimating victimization rates and public perceptions in local areas at lower cost than by adding to the NCVS sample. The first attempted design, an address-based sample (ABS) with

AAPOR 70th Annual Conference

telephone administration of the core NCVS instruments, had a low response rate and it proved difficult reliably to replicate NCVS data collection and processing procedures. Building on areas of success from the first design, a second design is being tested using ABS and a household level mail questionnaire. The new design differs from the core NCVS by using a household informant instead of self-reports from all persons age 12 and over in the household, and has fewer questions. These features support a less detailed crime classification than the NCVS. We recently completed a national pretest of this approach to assess feasibility, using two versions of the mail questionnaire with different approaches to asking about victimizations. The CS also includes questions on police performance and neighborhood safety, which are not part of the core NCVS but might be very valuable for evaluations at the local level. Previous research has suggested that such questions may affect victimization reporting, so the pretest included a split-ballot experiment on the placement of these questions, either before or after the victimization questions. The paper reports the results of the pretest, findings with regard to the feasibility of this CS approach, results of the two experiments, and lessons learned for the next phase of testing.

Building Probability Based Web Panels

Establishing the Probability-Based American Trends Panel

Chintan Turakhia, *Abt SRB*.

Nick Bertoni, *Abt SRBI*

Molly Caldaro, *Abt SRBI*

Charles DiSogra, *Abt SRBI*

Scott Keeter, *Pew Research Center*

Kyley McGeeney, *Pew Research Center*

Growing Internet access in the population plus the increasing prevalence of smartphones and other Web-accessible mobile devices are having a profound effect on survey research. In light of the ever-increasing costs for doing traditional telephone surveys, these trends in wider Internet access and rapid shifts in communication technology have accelerated the emergence and attraction of Web-based data collection. By comparison with multiple cross-sectional samples, panels provide a more efficient way of doing rigorous survey research and controlling costs without sacrificing the validity of generalizable, unbiased results. Internet access is not yet universal, of course, with an estimated 11% of U.S. adults still “off the grid.” And so any nationally-representative panel will need to account for the non-web population. Yet, the bulk of data collection in a panel can now be handled online. This paper presents the methods used to create the new, probability-based American Trends Panel (ATP) for Pew Research Center. This is a nationally representative panel of approximately 4,300 adults. The experience in panel building using a dual-frame telephone survey add-on recruitment module will be presented. The recruitment strategy for this effort, its implementation challenges and successful solutions are discussed. An evaluation of panel retention across multiple waves of surveying the panel is presented. We explore the demographic differences of panelists who are regular participants vs. those who are more occasional and those who joined but declined to participate. Methods for maximizing survey participation are presented. We also show survey completion rates by device type and mode since non-Internet enabled members (and those who chose not to do Web questionnaire surveys) are interviewed in a mail mode.

AAPOR 70th Annual Conference

Advance Postcard Mailing Improves Web Panel Survey Participation

Charles DiSogra, *Abt SRBI*

Kyley McGeeney, *Pew Research Center*

Scott Keeter, *Pew Research Center*

Andrew Burkey, *Abt SRBI*

Nick Bertoni, *Abt SRBI*

Molly Caldaro, *Abt SRBI*

Much research has been done on using pre-notification letters or postcards to improve survey response among de novo contacts. Those gains are generally 4-6 percentage points in mail surveys and thus a recommended part of the tailored design method described by Dillman. The reasoning is thought to be that people prefer to know that they are to be part of something important and can look for it to occur soon in the mail. It is also possible that it contributes to the main mail survey envelope actually being opened. On the other hand, members of a pre-recruited Web panel expect to do surveys on some periodic basis on the Web. A part of the panel protocol is for notifications to occur via email. When a survey is ready, an email invitation is sent with an active link to the survey questionnaire embedded in the text. This is and has been standard procedure for virtually all extant Web panels, including the recently created probability-based American Trends Panel (ATP) operated by Abt SRBI for the Pew Research Center. The relatively new ATP had several surveys administered in 2014 using email notifications. Persistent non-responders, however, were still present. In the event that email invitations were possibly not being seen, an advance postcard was mailed to panelists as an experiment for one survey. Panel members with a mailing address were randomized into two groups. About half (n=2,080) were mailed the advance postcard announcing their upcoming survey plus sent their usual email invitation. The other half (n=2,063) were sent the email invitation only. The group receiving the postcard had a 5.3 percentage point higher participation rate for the Web survey ($p<.001$). A comparison of demographic characteristics plus survey findings for responders and non-responders is shown to demonstrate any observed bias between the two groups.

Participation Effects in Panel Surveys: Evidence From Two Randomized Experiments

Sebastian Lundmark, *University of Gothenburg*

Mikael Gilljam, *University of Gothenburg*

Comparing panel samples and refreshment samples, previous studies have found significant participation effects on people's knowledge. Participation in previous panel waves tends to produce more knowledgeable respondents, but only minor effects have been found when studying people's beliefs, attitudes and voting intentions (Das, Toepel, and Soest 2011). However, most of these studies have used only two panel waves, and none of them have used a randomized experiment design. This study rectifies these shortcomings by using two panel studies with randomized experimental designs. With this more ambitious approach, we are able to study panel participation effects on political attitude questions, both in the short term, during high intensive panel participation (five surveys during four weeks), and in the long term, during low intensive panel participation (seven surveys during two years) in both cases with randomized gaps. More specifically, the design consists of one group of respondents receiving a set of panel questions in all the waves, and one group receiving the same set of panel questions in only fewer panel waves (e.g. receiving two gaps). Preliminary results show that

AAPOR 70th Annual Conference

some panel participation effect exists in the high intensive short term panel study but not in the low intensive panel study.

Web Survey Invitations: Design Features to Improve Response Rates

Jon Hughes, *Gallup*

Jenny Marlar, *Gallup*

The most important part of a web survey is arguably the email invitation. The invitation sets the stage for the rest of the survey, and allows the respondent to judge the costs and benefits of participation. The content of the invitation can also determine if the invitation gets delivered or blocked by email providers, and whether it lands in the recipient's inbox or spam folder. As web surveys increase in popularity but continue to have the lowest response rates of all modes, it is important to understand how the features of the email invitation impact non-response. Many features of the web invitation have been tested in past studies, such as subject line, personalization, contact and confidentiality information, topic, sponsorship, frequency of contact, incentives, and URL links. However, results have been mixed, and many studies were conducted in academic settings with unknown applicability to general population settings. Further, many studies have tested one feature versus another, but have not tested the combination of multiple features of the email invitation. The present study expands on the current body of knowledge by simultaneously testing multiple features of the survey invitation. These features include subject line, incentive, HTML versus plain text, sponsorship, personalization, and frequency of contact. The invitations were sent for two different types of general population studies. The first is the Gallup Panel, a probability-based panel of 55,000 members who receive survey invitations on a regular basis. The other is a customer satisfaction survey, in which respondents have a customer relationship with a business but are not expecting a survey invitation. The findings will include results on bounce backs, SPAM flags and spamminess scores, open rates, click through rates, and final response rates. Recommendations are also made for features that improve deliverability, open-rates, click through rates, and ultimately response rates.

Interviewers, Interviewing and Data Quality

Understanding Paralinguistic and Linguistic Strategies in Research Interviews

Casey Langer Tesfaye, *The Nielsen Company*

Darin Harm, *The Nielsen Company*

In a standardized research interview the comparability of responses is paramount, and interviewers have very little freedom to tailor their script as the interview progresses. However, even under these constraints the best interviewers seem to be able to make a standardized interview feel less artificial. A more conversational feel can be achieved by adapting the script, but these adaptations can compromise the quality of the responses. A more natural feel can also be achieved without changing the script at all, through paralinguistic strategies. For this paper we will code and use audio recordings from a call center to explore the linguistic and paralinguistic tools that interviewers use in order to make scripts appear more conversational. The linguistic strategies we will discuss include personal recall and script tailoring. The paralinguistic tools that we will consider include pacing, tone, register, volume, creaky voice and laughter. We will explore specific uses of each and the ways in which the use of these features influences the research conversation. We will also discuss the systematic use of these features

AAPOR 70th Annual Conference

to provide cues that signal shifts from reading to asking and from asking to answering, thus facilitating the question and answer process. The goal of this research is to develop a more systematic, data driven, and practical guide to better train telephone interviewers.

Accessing Quality of Interviewer Observations in Measuring Subjective Questions

Mengmeng Zhang, *American Institutes for Research*

Lindsay Ryan, *University of Michigan*

Jacqui Smith, *University of Michigan*

Interviewer observations are used to evaluate survey measurement error and adjust for post-survey non-response bias. As a result, it is important to examine the accuracy of the interviewer observations in order to avoid biased adjustments. Previous literature has examined the accuracy of interviewer observations by comparing observations on objective variables with record data, yet little research has focused on the quality and utility of interviewer observations for measuring subjective questions such as general well-being or cognitive functioning. This paper utilizes data from a national longitudinal health panel survey that targets older adults to evaluate the accuracy of interviewer observations of both objective and subject health-related variables. We hypothesize that interviewer observations are more accurate for objective variables than for subjective variables. To determine this, we first utilize multilevel modeling to examine the extent of disagreement between interviewer observations and respondents' responses in terms of several relatively objective (such as eyesight, hearing, and functional limitation) and subjective (such as memory, self-rated health, and emotions) health-related variables. Health record data of the respondents are used to assess the accuracy of the respondent reports for subjective questions (such as memory score, animal counts, number of diseases, and other medical records). Second, we explore the factors influencing the accuracy of interviewer observations for subjective well-being questions, accounting for respondent characteristics, emotions during interviews, the level of survey cooperation, and other auxiliary data. The findings can provide insights to future research regarding the usability and thresholds of using interviewer observations to evaluate measurement error and adjust for non-response bias and for subjective questions.

A New Method for the Analysis of Interviewer Variance, With an Empirical Application

Patrick Sturgis, *University of Southampton*

Ian Brunton-Smith, *University of Surrey*

George Leckie, *University of Bristol*

It is well-known that interviewers tend to inflate standard errors, relative to self-completion modes, through the ways in which they administer survey questions (O' Muirheartaigh and Campanelli, 1994). When a question is asked across respondents in the same idiosyncratic manner by the same interviewer, a within-interviewer dependency is created. For survey analysis, this manifests in the form of greater variance in the population estimator. Traditional methods for the identification of interviewer variance in observational studies rely on estimation of random effects in multi-level models, where the random effect is specified at the interviewer level (e.g. Brunton-Smith and Sturgis 2012). This approach suffers from a number of analytical limitations, notably a lack of flexibility in modeling the effect of interviewer-level characteristics on response variance. In this paper we set out a new method for the analysis of interviewer variance; the mixed-effects location scale model extends the random part of a standard 2-level mixed effects model by introducing random effects on the level 1 variance (Leckie, 2014). This

AAPOR 70th Annual Conference

enables response variance to be re-parameterized in a way that affords a more flexible and causally focused assessment of the factors associated with interviewer-induced response variability. We apply this approach to data from wave 3 of the UK Household Longitudinal Survey (UKHLS), which we link to a diverse range of interviewer characteristics measured in an independent survey of interviewers. We demonstrate how the modeling strategy can be used to identify individual-level characteristics which are associated, conditionally, with larger interviewer variance components. We also show how the approach might be used to improve survey quality by identifying interviewers, and interviewer characteristics, that are associated with more variable survey responses.

Do Interviewer Effects Matter: Evidence from European Social Survey

Ashley Amaya, *JPSM at University of Maryland*

Delancey Gustin, *University of Maryland*

Herschelle Lisette Sanders, *University of Maryland*

Ji Qi, *University of Michigan*

Interviewers often introduce variance and bias in survey estimates. Respondents alter their answers to be more socially desirable so as to avoid judgment by the interviewer. Alternatively, interviewers may deviate from the standardized question wording which alters the way in which the respondent perceives the question and how he/she answers it. This potential for increase variance and/or bias has resulted in a call by survey methodologists for researchers to account for interviewer effects in their analyses. Despite the call for their inclusion, researchers rarely account for interviewer effects in their analyses. While simulations and mathematical proofs have demonstrated the potential for spurious inference when interviewer effects are ignored, it is not clear that interviewer effects are commonly large enough to alter the final conclusions drawn from the data. The purpose of this paper is to determine whether or not their exclusion matters. Would researchers draw the same conclusions from their data regardless of whether or not they control for interviewer effects? We use an example in the political science literature. Using 2012 European Social Survey (ESS) data for Belgium, we retested the hypothesis that individuals with higher perceptions of societal fairness will have more positive attitudes toward immigrants. Previous research has found support for this hypothesis, although interviewer effects were not included in the analyses. To retest the hypothesis, we first evaluated a series of mixed models to account for interviewer effects. We then ran a simple linear regression to recreate the analyses previously used to test this hypothesis. Finally, we evaluated their hypothesis under both types of models to determine whether our conclusion would have changed.

Panel: Social Media Data Mining: Staying on the Cutting Edge

Panel: Social Media Data Mining: Staying on the Cutting Edge

David A. Schweidel, *Goizueta Business School, Emory University*

For the last few years, public opinion researchers and survey methodologists have been exploring how to take advantage of the wealth of data on social media. Last year, the AAPOR Task Force on Emerging Technologies put out a report on Social Media and Public Opinion Research: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research. This panel builds on that report bringing in researchers from other disciplines, like

AAPOR 70th Annual Conference

communications, evaluation and marketing to help survey researchers keep on top of cutting edge science in the area of social media data mining.

Assessing Brand Perceptions with Social Media

David A. Schweidel, *Goizueta Business School, Emory University*

Social media conversations offer a rich source of insights. In contrast to traditional tracking studies implemented by brands, social media can offer a more timely and economical means of learning about consumer perceptions. But, can they be trusted? Unlike survey-based methods for learning about consumers, turning to social media conversations entails that researchers relinquish considerable control. We will discuss the concerns that have been raised with regards to using social media conversations as a means of supplementing traditional marketing research techniques, including the construction of the sample and construction of specific topics of interest. While there are limitations that have been identified in turning to social media conversations, there are potential benefits. We will discuss the marketing insights that can be derived from the analysis of social media data, including assessing perceptions toward brands and products, and the identification of those attributes on which the brand performs strongly or weakly. We will discuss biases that exist in social media that need to be taken into account to derive these inferences. In addition to discussing the use of social media as a means of understanding consumer perception about an individual brand, we will also discuss how it may be used to assess brand perceptions relative to their competitors.

Digital Research on Climate Change: Turning to Social Media

Jason Boxt, *Glover Park Group*

Colleen Campbell, *Glover Park Group*

The United Nation's Intergovernmental Panel on Climate Change (IPCC) provides comprehensive assessments of current scientific, technical and socio-economic information about the risk of climate change, its potential environmental and socio-economic consequences, and options for adapting to these consequences or mitigating the effects. Because the IPCC report is always the focus of fervent opposition and expressions of support and drives discussion in dozens of countries, the Glover Park Group (GPG) was approached by the United Nations Foundation (UNF) to help communicate around its release. Social media and the explosion of publicly available user content have driven researchers to look to new and innovative ways of conducting research in order to understand how people think, talk about, and act upon issues. Traditional opinion research, using both quantitative and qualitative methods, must look beyond the standard focus group or telephone poll to help clients achieve their communications objectives. For this reason, GPG conducted four pieces of research: a comprehensive opinion audit and analysis, a global digital environmental scan across 14 countries and 13 languages, online surveys in four countries and a global source map illustrating the strength of mentions and backlinks. For the digital scan, GPG developed a list of 1,796 keywords across nine topic categories to collect data in English and native languages in 14 countries. More than 10 million conversations were collected from January 1 –December 31, 2012, from online news, blogs, message boards, microblogs, social networks, review sites and multimedia sites. Analysts were randomly assigned a statistical sampling of data from the English and the native language results, ensuring that the data would be representative of the most predominant discussions happening in each country. Fascinatingly, the public perception found in the open source digital media scan reflected the results found in the online surveys. This presentation will highlight those findings.

Development of Age-Prediction Algorithms for Twitter Followers

Antonio Morgan-Lopez, *RTI International*

Researchers are increasingly mining social media data for insights into consumer beliefs and behaviors, but knowing who these data represent is challenging. Demographics on social media users are typically proprietary and, if available, are in aggregate form. While individual-level demographics could be determined through surveys, this approach would be cost prohibitive for large scale social media analysis. A more promising approach, and the goal of this analysis, was to develop algorithms to predict the age of social media users based on publicly available content of their posts. We created a labeled dataset of Twitter users based on birthday announcements (e.g. "happy 18th birthday to me!"). We identified 1279 Twitter users ages 13-24 and pulled the last 200 Tweets for each user. We restricted the analysis to words that appeared at least 20 times across each individual's 200 Tweets, with 6641 such words being retained. The sample was split into a calibration sample (N=700) and confirmation sample (N=579). With the calibration sample we estimated separate, single predictor OLS regression models where each of the 6641 word predicted age. Words that were statistically significant at alpha 0.05 level were then used in a multiple predictor OLS model with the confirmation sample, and the model-predicted ages were compared to "actual" labeled ages based on birthday announcement. We achieved good model fit ($R^2 = 0.682$) and 89.7% accuracy in predicting youth and 75.5% accuracy in predicting adults. Words that were predictive of age centered around work, college, and travel for adults and school subjects, references to Tweeting, and text shorthand for youth. The analysis was repeated with words that were predictive of age from a survey sample of Facebook users (Schwartz et al, 2013) and results were comparable. Our findings suggest that mining Tweets may be a promising approach for predicting age of Twitter users.

Topic Discovery in Text-Driven Social Science Research

Philip Resnik, *University of Maryland*

Social media represent a tremendous opportunity for improved understanding of individuals and populations: conversations that used to take place outside the reach of scientific inquiry, at the office water cooler, over a dinner table, or at happy hour after work, are now taking place in public settings that make data collection and analysis possible. The resulting datasets, though, present challenges for traditional content analysis methods. Manual coding approaches can't scale up, and computational analyses based on words alone are limited in their ability to capture higher level concepts and themes in the face of huge variability in how language is used. This talk will discuss recent work that complements more traditional approaches by performing large-scale computational analysis above the word level. "Topic modeling" is a family of computational models that exposes latent structure in text collections, making it possible to discover underlying topical patterns without either presupposing a codebook or relying on individual words. Although they are fundamentally driven by noisy, unstructured text, extended versions of basic topical models make it possible to take advantage of "clean" information when available, e.g., in the form of human domain expertise, manual coding, survey responses, or other metadata, in order to discover more relevant structure and create predictive models.

Sunday, May 17
10:15 a.m. – 11:45 a.m.
Concurrent Session K

Mini-Conference: Non-Probability Samples in Election Surveys and Beyond

Reaching Wider, Going Deeper: Incorporating Sample Source Variation and Other Considerations into MRP Adjustments of Polling Estimates for Blended River Samples

Robert A. Petrin, *Ipsos Public Affairs*

Neale A. El-Dash, *Sleek Data*

Multilevel regression with post stratification (MRP), has been proposed as a principled method for generating population-valid inferences using nonprobability samples (Gelman and Hill, 2006; Ghitza and Gelman, 2013; Wang et al., 2014). MRP is an extension of small area estimation (e.g., Rao, 2003; Carlin and Louis, 2008) which requires first obtaining population targets based on the “deep stratification” of a number of key respondent attributes. Bayesian methods are then used to generate estimates for the corresponding sample stratification cells (since for the sample a great number of these cells are generally either highly sparse or empty). Finally, traditional post stratification methods (Heeringa et al., 2010; Kish, 1995; Korn & Graubard, 1999) are applied to obtain parameter estimates and inferences. To date the performance of MRP has been examined using data from a single nonprobability samples and applying conventional Bayesian models. In this paper we evaluate ways of extending MRP by considering adaptations that have the potential to improve its performance when used with online blended river samples. Among these adaptations are: modeling between-source variation in the blended sample; using adaptive compound-process priors, placing linear constraints on priors, and applying priors derived from relevant market sources. The data used for the empirical portion of the paper include two separate continuous tracking poll time series: the Ipsos-Reuters continuous tracking poll of Obama approval March 2011 - November 2014, and Ipsos-Reuters 2014 pre-Midterm Election polling for selected battleground states. Results from a simulation study are also presented which compare MRP to traditional probability sampling procedures. The performance of each of the proposed extensions of MRP are evaluated using Jackman’s (2005) method for assessing bias in political polls. The paper concludes with caveats and recommendations for extending MRP to political, public opinion, and public health research.

Self-Reported Voting Patterns on the Day of the Election

Zachary H. Lewis, *Ipsos Public Affairs*

Alan Roshwalb, *Ipsos Public Affairs*

Over the past two elections—2012 and 2014, Ipsos has polled its online panel on the day of the election asking panelists to report their vote. This survey captures voting behavior and attitudes among both early voters and election-day voters. In this paper, we examine reported voting patterns from 10 am on Election Day through the evening to identify trends during the election. These patterns are examined to understand their usefulness in projecting the outcomes of national and state-wide elections. Several models are used to assist in forecasting including

AAPOR 70th Annual Conference

Bayesian shrinkage estimators. The paper concludes with caveats and recommendations for this approach's possible use in for identifying trends in voting behaviors and predicting outcomes on Election Day.

Acing the Midterms: A Unique Approach to Pre-Election Polling

Jon Cohen, *SurveyMonkey*

Sarah Cho, *SurveyMonkey*

Noble Kuriakose, *SurveyMonkey*

As an experiment in pre-election polling, SurveyMonkey polled nearly 200,000 people in the U.S. from Oct. 3 to Nov. 6, 2014, recruiting off the nearly three million user-created surveys running each day on the platform. Using all the responses among adults with only standard age, race, sex, and education adjustments, the pre-election estimates had the right winner in all 36 U.S. Senate races, and 33 of the 36 gubernatorial contests in the 2014 midterm elections. In this paper we explore the substantive findings, methodology, click-through and completion rates, and generalize about successful, future approaches to pre-election survey work.

Weighting to Scale: The Nature and Measurement of Selection Effects in Online Sample

Yannick Dufresne, *Université Laval*

Charles Tessier, *Université Laval*

Clifton van der Linden, *Vox Pop Lab*

Mounting evidence regarding the shortcomings associated with RDD sampling have eroded the gold standard of telephone surveys in an era where landlines and telecommunication are on a sharp decline. Prompted by a shift in the communications landscape, public opinion researchers have turned to emergent information communication technologies as a corrective to representative sampling. However, the selection effects inherent to many of these collection instruments require increasingly complex statistical treatments in order to ensure robust, representative inferences. Election forecasting offers social and statistical scientists a unique opportunity to test for the selection effects present in online sampling techniques given that elections serve as an empirical benchmark by which to validate competing statistical treatments of online panel. This paper leverages data collected by Vox Pop Labs to assess the accuracy and robustness of various online panel weighting techniques. The Vox Pop Labs data contains more than 5 million voter observations across a dozen elections in Canada, the United States, Australia and New Zealand. The respondents variously profiled by different attitudinal and socio-demographic traits. A subset of the data tracks the same respondents over multiple election campaigns. The Vox Pop Labs data will be used to test whether the selection effects of online sample are constant over time and, if so, whether the sample weights applied to a recurring observation within the dataset are constant over time. The findings from this study will have significant implication for public opinion research using Big Data, particularly with a view to making representative inferences from self-selected sample.

AAPOR 70th Annual Conference

Meta-Analysis of Online Panel and Non-Panel Sampling: Electoral and Non-Electoral Behavior Metrics

Julia Clark, *Ipsos Public Affairs*

Clifford Young, *Ipsos Public Affairs*

Robert Petrin, *Ipsos Public Affairs*

The survey research industry is in an era of transition and flux: 'traditional' methodologies (now defined as live interviewer landline phone and cellular phone) are waning in prevalence – if not importance – while newer methodologies (online, IVR, blended) are gaining prominence. The impact of this on measurement accuracy is still being explored, with new information emerging all the time. To further this body of knowledge specifically within online work, we analyze behavior measurement accuracy of the two main sampling approaches for internet-based research: panel only and non-panel (River and other approaches) sampling. Our paper will comprise a meta-analysis of nonprobability online sampling, looking across over 400 data points, at election and non-election behaviors as validating points. Behavior measurement accuracy will be assessed against external benchmarks from market-accepted data sources such as the U.S. Census, Current Population Survey (CPS), and the General Social Survey (GSS). The analysis will focus specifically on differences in accuracy between sample sources or blended compositions, as well as which socio-demographic subgroups are most accurate across a range of different validation measures.

Applications from Market Research to the Survey World

Digging Deeper: Exploring Consumers' Subconscious Perceptions in Survey Research

Megan Peitz, *Gongos, Inc.*

Joe Cardador, *Gongos, Inc*

Much of the research on brand associations involves asking consumers explicitly about their attitudes toward particular brands. It is well established that people are not always rational in their decision-making (e.g., Kahneman, 2013) and that the design of survey questions and ratings scales can impact survey results (e.g., Krosnick & Fabrigar, 1997). This suggests that there may be implicit beliefs motivating behavior that are difficult to capture using traditional survey approaches. As researchers, how do we understand consumers' non-conscious attitudes about specific brands or products and how can we use this information to better predict consumer behavior? One tool available to measure these underlying attitudes is the Implicit Association Test (IAT). We asked consumers to sort adjectives and pictures into buckets for two different home improvement retailers as quickly as they could. The errors they made and the speed at which they were able to complete the exercise determined their underlying brand associations. Four unique tests were designed using the Inquisit® platform from Millisecond Software and results were scored using a script developed in R (which can be shared) that applies a version of the algorithm created by Greenwald, Nosek, and Banaji (2003). This algorithm uncovers the relative association of a brand with a particular adjective for each individual. Results showed that there are adjectives that respondents implicitly associate with one brand over another, offering advantages in marketing and communication. Explicit attitudes expressed by survey respondents did not always align with their implicit associations. The link between explicit ratings, implicit associations, and consumers' shopping behavior at the brands

AAPOR 70th Annual Conference

studied is examined. Initial findings suggest that implicit methods can contribute to a deeper understanding of consumer brand preferences and choices.

Adapting Conjoint Techniques to the CATI Environment

Edward Paul Johnson, *Survey Sampling International*

Pete Booth, *Infosurv*

Choice based conjoint, with its ability to closely mimic realistic decision making scenarios and produce robust part-worth utility scores, has become one of the most widely used methods of quantifying preference in modern research. However, one of the important drawbacks of conjoint is that it relies heavily upon the visual real estate to allow respondents to compare products. Unfortunately sometimes the population of interest is too hard to reach online and interviews need to be done over the phone. We explore an option of a Simple Conjoint (proposed by Jordan Louviere as Best Worst Case 2) that can give the results of a conjoint but keep the tasks simple enough to complete over the phone. We test this Simple Conjoint in a real life scenario: in this case with a shipping company. We completed a little over 650 telephone surveys in the Texas region with a hard to reach business audience. The shipping options we presented them had four attributes with three levels each. Over the phone we describe one product from the balanced design at a time and the respondent tells us which attribute of the product is most appealing and least appealing to them. Products are never compared against each other because it is too hard for a respondent to keep more than one product in their mind at a time. We show how the principals of the simple conjoint were able to achieve a 98% cooperation rate over the telephone interview on the conjoint portion of the survey while still obtaining the part-worth utilities needed for the analysis. We also discuss the limitations of the simple conjoint including the inability to measure interaction.

Audience Segmentation to Support Consumer Engagement in Using Healthcare Benefits

Frank Funderburk, *Centers for Medicare and Medicaid Services*

Diane Field, *Centers for Medicare and Medicaid Services*

Clarese Astrin, *Centers for Medicare and Medicaid Services*

Audience segmentation can play a key role in developing effective social marketing campaigns. This presentation provides an overview of the development of such a framework via a mixed-methods approach that combines qualitative formative research and quantitative audience segmentation analysis, including demographic, geographic, and psychographic perspectives. We show how this framework can be applied to improve communication with audiences that will be participating in various agency programs, including the emerging health insurance marketplace. We used a mixed methods research strategy for this work. A series of qualitative and quantitative studies set the stage for the application of a latent class segmentation methodology that suggested a six-segment solution. This initial framework was validated in a field study (N = 4,200) of consumer response to prevention messages. Subsequent analysis explored differences between the insured and uninsured on dimensions related to the segmentation framework. We achieved high levels of both internal and external validity by building on the strengths of diverse research strategies. Our work has found that a six-segment framework developed for general healthcare audience segmentation worked well with the target audience for the emerging health insurance marketplace. Over 92% of the uninsured fell into three of the segments: Healthy & Young, Sick, Active & Worried, and Passive & Skeptical. These segments shared some common characteristics related largely to their insurance status, but also differed from one another in systematic ways that suggest distinct attitudinal and

perceptual barriers that required attention to achieve effective engagement. The iterative process adopted in this mixed methods approach has accelerated the development of core knowledge of the target populations, produced measurable improvements in outreach and educational effectiveness, and increased the return on Agency investments in communication.

Methodological Briefs: Sampling and Frame Building

Probabilistic Record Linkage to the National Plan and Provider Enumeration System for Data Recovery and Validation in Physician Sampling Frames

Akash A. Desai, *American Institutes for Research*

Grace Wang, *American Institutes for Research*

Sarah Ng, *Amgen*

Physician sampling frames often contain erroneous or missing identifier and/or contact data. Erroneous or missing data can lead to under-coverage in a sampling frame and potential bias of estimates (Degaetano, 2013). Additionally, non-response due to erroneous contact data is costly. As one potential solution to these threats to data quality, we demonstrate the use of probabilistic record linkage to the National Plan and Provider Enumeration System (NPPES) as a tool to recover and validate physician identifier and contact data. The Health Insurance Portability and Accountability Act (HIPAA) provided for the establishment of the NPPES and mandated the adoption of standard unique identifiers, called National Provider Identifiers (NPIs), for healthcare providers. Healthcare providers must use their NPI in administrative and financial transactions covered under HIPAA, and are widely adopted. For this study, we used the Office of the National Coordinator's Regional Extension Center Customer Relationship Manager (CRM), which is an administrative dataset containing contact information for clients of a national technical assistance program. We used the Stata add-on reclink on provider names and addresses to recover missing NPIs from the CRM file. We examined the precision, defined as true matches over test outcome matches, along the distribution of match probabilities produced by the program using visual inspection. We recovered 46 percent of NPIs, seemingly without error, for physicians who lacked them in the CRM using a discrimination threshold that allowed for 5 percent record linkage error. Thresholds that tolerate higher record linkage error can lead to marginally higher recovery with error. We describe considerations for weighting, validation of geocoding data, and surveying. We conclude that probabilistic matching on the NPPES database is an effective and low-cost first step for data recovery and validation when constructing physician sampling frames and geocoding.

Comparing Surveys Based on RDD and ABS Samples Draw to Represent the Same Populations: Are There Demographic and Health Differences?

David R. Johnson, *Pennsylvania State University*

Yunfeng Shi, *Pennsylvania State University*

Donald S. Miller, *Pennsylvania State University*

Low response rates and the growth of cell phone use have reduced efficiency and coverage in RDD telephone surveys necessitating alternative strategies for population based surveys. Use of address based sampling (ABS) to select the study participants may be a viable alternative but direct comparisons of RDD and ABS based surveys drawn from the same population are rare. This paper compares surveys conducted three years apart (2010 – 2013) in three metropolitan areas. In the first round an RDD-based design was used to sample both landline and cell phone

AAPOR 70th Annual Conference

numbers. In the second round the same survey was repeated but it was based on an ABS sampling design. In the ABS design, when telephone numbers were available for the sampled addresses, these called to complete the survey by telephone. Those without matching telephone numbers were contacted by mail and were asked to provide their telephone number so they could be interviewed. A mail survey was not used so comparability with the first round could be maintained. There were 3,238 respondents in 2010 and 2,976 in 2013. The AAPOR 4 response rates was 37.0% in the RDD and 8.4% in the ABS based surveys. Based on ACS data, little change over three years in demographic distributions in the populations was found, allowing us to compare the distributions on basic demographic variables in the two survey types. We also examined differences in the variability of the post-stratification weights. Preliminary results show significant differences between the RDD and ABS surveys in demographic distributions and the variability of the weights, raising concerns about the comparability of the two approaches. Additional multivariate analyses will be conducted to explore these differences and their effects on estimates of self-rated health and chronic conditions.

Use of Chain Referral Sampling to Build a Panel of Latino Families

Christine Cowles, *Abt SRBI*

Mary Haan, *University of California, San Francisco*

Allison Aiello, *University of North Carolina, Chapel Hill*

Snowball or chain referral sampling is often used by survey researchers to study populations that are rare or difficult to access. It is also appropriate for identifying family and other social networks in which survey participants belong. The Sacramento Area Latino Study on Aging (SALSA Study) conducted between 1998 and 2007 tracked the incidence of physical and cognitive impairment as well as dementia and cardiovascular diseases in a cohort (n=1,798) of elderly Latinos in the Sacramento California region. As part of the SALSA Study, contact information for the respondents' children was collected. The Niños Lifestyle and Diabetes Study conducted in 2013 (Wave 1) and 2014 (Wave 2) examined the relationship between lifestyle, acculturation, health behaviors, family history and risk of diabetes across generations among a cohort of Latino families. The baseline panel of Niños Study subjects (n=534) was comprised of the biological offspring of SALSA Study participants. In each wave of data collection Niños Study participants were asked to refer biological relatives aged 18 years and older into the study panel. Results indicate that 38% percent of the Wave 1 respondents and 35% of the Wave 2 respondents referred at least one family member. This paper will describe how the chain referral process was implemented and how challenges were addressed. It will report the level of referral redundancy and saturation point. We will also report the staff and field period time requirements for chain referral sampling with this population and discuss our assessment of the financial and time costs of chain referral sampling.

Are Women Less Likely to Answer Cell Phones?

Sarah Dipko, *Westat*

Darby Steiger, *Westat*

David Cantor, *Westat*

A feasibility test conducted in May 2014 with a dual-frame cell phone and landline sample yielded a surprising result for the cell phone sample. The feasibility test was 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 gender. However the yield from this test was

AAPOR 70th Annual Conference

substantially lower than expected, with an observed eligibility rate of just 23 percent. Age eligibility was close to expectations at 78 percent; however gender eligibility was very low at just 30 percent. An examination of refusal cases indicated an even balance by gender, and it was hypothesized that the “missing women” were among the cases that never answered the phone across seven call attempts. A second cell phone sample feasibility test was conducted in August 2014. Questions were added for interviewers to code upon reaching voice mail, to determine the likely gender of the person connected to the phone number. This second test also achieved a lower than expected eligibility rate of just 27 percent, with 74 percent of phone numbers reaching adults, and 36 percent of those adults being female. Examination of the new gender of voice mail questions indicated that among cases where gender could be ascertained, 78 percent were coded as reaching females. Recent research describes the phenomenon of cell phone samples under-representing younger women and over-representing older women (Ben-Porath and Sherr, 2014). Our feasibility tests indicate a different problem – women in general are significantly less likely than males to answer their cell phones, regardless of age. The age distribution of our female respondents was close to the population distribution, but the representation of women overall was low.

Comparing Eligibility Rates and Demographic Characteristics Across Multiple Online Recruitment Methods in a Smoking Cessation Study

Derick Brown, *RTI International*

Linda Squiers, *RTI International*

Jill Dever, *RTI International*

Janice Tzeng, *RTI International*

Brian Southwell, *RTI International*

Suzanne Dolina, *RTI International*

Sidney Holt, *George Washington University*

Amy Sanders, *ICF Interactive*

Erik Augustson, *National Institutes of Health*

There are a growing number of surveys recruiting and collecting data from participants solely through online methods. As such, the issue of recruitment methods for online surveys is an important, yet relatively unexplored area for research. The purpose of this presentation is to share the results from an innovative study, NCI's QuitTXT Smoking Cessation Evaluation (QuitTXT), that conducted online recruitment and data collection without any traditional sampling methods. The target population of QuitTXT was young adults, ages 18-29, who currently smoke cigarettes and want to quit. Recruitment tactics included online banner advertisements, search engine optimization, and online marketing vendors (e.g., Google, Yahoo, Facebook) to attract eligible individuals to the survey. We compared study eligibility rates and demographic proportions by recruitment tactics to assess effectiveness. The findings will help to inform future studies that plan to use online-only surveys by developing more efficient recruitment strategies, while simultaneously controlling the balance across the demographic and participant characteristics of interest.

To Re-Mail Or Not to Re-Mail: Evaluating Occupancy Status in an Address-Based Household Mail Survey

Cameron Brook McPhee, *American Institutes for Research*

Michelle Cantave, *American Institutes for Research*

Mark Masterton, *American Institutes for Research*

Surveys using address based sampling (ABS) methods require precise tracking of housing unit eligibility. The occupancy status of the housing unit is a key aspect when determining unit eligibility, response rates, and non-response adjustments. Additionally, mail-based studies typically utilize multiple contacts, sometimes using increasingly expensive contact strategies (e.g., special delivery mail). These additional contacts are unnecessary if an address is confirmed to be unoccupied, nonresidential, or nonexistent. However, because it is often difficult to confirm the occupancy status of sampled addresses, organizations often retain all housing units in subsequent mailings even if early attempts indicate the unit is unoccupied. The U.S. Postal Service and special delivery services (e.g., FedEx) may provide different information about a unit's occupancy status. Further, the extent to which occupancy information collected by USPS is comparable to information collected by special delivery services and whether it is beneficial to use special delivery mailings in follow-up contact attempts if USPS has indicated the unit is unoccupied remains unknown. This analysis addresses these questions by analyzing data from the 2014 National Household Education Surveys Feasibility Study (NHES-FS). The NHES-FS is an ABS mail survey sent to an initial sample of 60,000 addresses across the U.S. The first two mailings were sent using USPS First Class mail and the third contact was sent via FedEx. Occupancy information will be compared across delivery services, as well as to occupancy status indicated on the sampling frame. These analyses will be conducted by region and urbanicity. Response status will be analyzed to determine the impact on response rates of re-mailing to addresses identified as unoccupied. This analysis will evaluate the effectiveness of this strategy and determine whether a refined methodology that accounts for frame information, location, and postal returns could be used to more accurately identify unoccupied housing units and reduce expenditures.

Automated SMS Text Messaging as a Tool in Public Opinion Research

Nina DePena Hoe, *Institute for Survey Research - Temple University*

Heidi Grunwald, *Institute for Survey Research - Temple University*

Keisha Miles, *Institute for Survey Research - Temple University*

Speaking to a need for innovative methods in data collection, this methodological brief demonstrates the utilization of web-based, automated SMS text messages in public opinion research in an attempt to cost-effectively reach more diverse samples of citizens, through their cell phones. The Textizen™ web-based SMS text message platform allows for the large-scale administration of short, text-based surveys. Of particular interest to public opinion researchers, some traditionally harder-to-reach populations, such as low-income and minority populations, have been quick to adopt mobile phone technologies, and use text-messaging services a higher rate than others. This presentation will reference a recent test of SMS text messages as a “primary” mode of data collection – both in recruiting participants and administering a survey in a regional context. A sample of 1,000 mobile subscribers was contacted via a “cold text” asking them to participate in a 5-question survey regarding their opinion on a local park; a raffle of 10 iPads was offered as incentives. Non-respondents received follow-up phone calls to determine their reason(s) for not responding, such as a non-working number, the inability to receive text messages, non-interest in text-message survey participation, or invasion of privacy. Significant empirical results, presented in a table, include the rate of text message non-receipt, rate of text

AAPOR 70th Annual Conference

message receipt, frequency of reported barriers to responding (reasons for not responding), and demographics of respondents. Findings provide rich insight into the use for automated text messages in public opinion research in general, and have implications for particular local and global markets where the number of mobile subscribers and text message usage is high. Findings suggest automated SMS text messages are an effective way to measure public opinion across wide variety of citizens, and cost effective when compared to other methods.

Using Auxiliary Data to Increase Efficiency of Sampling Rental Units

Randal ZuWallack, *ICF International*

Joshua Brown, *ICF International*

Thomas Brassell, *ICF International*

Davia Spado, *ICF International*

The Department of Housing and Urban Development (HUD) annually publishes Fair Market Rents (FMR) for metropolitan areas and non-metropolitan counties. HUD primarily uses data from the American Community Survey (ACS) to determine the FMRs. For small counties, the ACS data is limited by small sample sizes and data recency. HUD uses adjustment factors, to account for these limitations, but adjustments for small areas are often based on larger geographic areas. Public Housing Authorities who believe that the ACS-based FMR does not reflect the rental market are allowed to submit alternative data to HUD for FMR calculations. Cost is a barrier in conducting local area rent surveys due to low incidence. In a recent local area rent survey sponsored by the Vermont State Housing Authority, we selected 10,000 addresses from three counties using a list updated according to the computerized delivery sequence file. Of the 3097 valid returns, 508 qualified for analysis. To investigate cost reductions, we conducted an experiment using two auxiliary pieces of information, a tenure flag appended to the addresses and the percentage of rental addresses in the address' geographic area. We found 95% accuracy of the tenure flag in predicting home ownership, which is similar to other research on the accuracy of this information. The exclusion of addresses flagged as "owned" will increase efficiency in reaching rented homes, but could have bias implications since 18% of rental returns were from addresses flagged as owned. We further explore whether the use of geographic rental density can improve on identifying higher rates of rental returns. We balance the cost benefits with bias and variance implications to design an optimal sampling strategy for sampling rental units.

Surveys on Science, Energy and Climate Change

The Influence of Political Ideology on Politicized Beliefs About Science

Jason B. Reineke, *Middle Tennessee State University*

Rebecca R. Donaway, *Middle Tennessee State University*

Scientists and policy makers alike have vested interests in understanding the public's support for and understanding of science (PUS). Researchers have looked to numerous predictors of belief regarding issue attitudes about various science topics, such as schema, science knowledge, religious belief, and trust in scientific authority. Some topics can be treated as issue knowledge, like the awareness of increased carbon levels in the atmosphere. However, in today's increasingly partisan political environment measuring attitudes about how to address contentious issues like climate change are more complex. We assert these questions, and their responses, are ultimately tapping into ideas about the role of government to regulate

AAPOR 70th Annual Conference

businesses and people and the impact humans have on the environment. Hindman's (2009, 2012) "belief gap" hypothesis suggests that, beliefs about politically contested issues can displace issue knowledge as the most important dependent variable, and political ideology can act as the best predictor of either knowledge or beliefs. This theory, applied to science issue attitudes would suggest that ideology, not science knowledge or any other variable, would be the best predictor of belief. Based on secondary analysis of GSS Survey data, this study explores the question of ideology's role in predicting politicized beliefs about climate change by proposing and testing a structural equation model. Specifically, the study examines ideology, science knowledge, and schema with regards to three related questions about how well scientists understand climate change, whether or not they should influence policies about climate change, and whether or not the government should regulate gas prices. The findings support the hypothesis and the implications for ideology's role in beliefs about science are discussed. Additionally, this study tests, and finds support for, Roos's (2012) separation of the National Science Foundation (NSF) science literacy scale, based on two specific questions, which measure ideas about religion rather than science.

Attitudes Toward Unconventional Energy Production: An Emerging Political Prism?

Erik P. Bucy, *Texas Tech University*

Melissa R. Gotlieb, *Texas Tech University*

Bryan McLaughlin, *Texas Tech University*

This paper reports the results of a national online panel survey conducted through SSI in summer and fall 2014 (N = 2,500 in wave 1, and approximately 1,800 in wave 2) designed to measure awareness and attitudes towards unconventional energy production technologies, including wind and solar energy, hydraulic fracturing ("fracking"), hydrogen, and grid energy storage. Respondents were asked how much they followed news about unconventional energy production as well as support for government policies that would increase various forms of production. They then assessed the perceived risk associated with each of these production technologies, in terms of the benefits outweighing the risks, or risks outweighing the benefits. Risk estimates and support for unconventional energy production are first analyzed geographically, by place of residence, and then from the perspective of value holding. Whether attitudes toward energy policy are supportive or oppositional relates to the extent to which respondents prioritize energy independence and economic opportunity on the one hand, or environmental risks and quality of life considerations on the other. Finally, we analyze strength of support for different forms of unconventional energy production as a predictor of political involvement in the 2014 midterm elections, testing the supposition that the "greener" one's energy outlook, the greater support for Democrats, while the more concerns are focused on economic matters, the greater support for Republican candidates. Results are discussed in light of the growing importance of attitudes toward alternative energy policy as an emerging political prism.

Climate Change Policy and Public Opinion in Canada

Keith Neuman, *The Environics Institute for Survey Research*

Climate change is a global challenge that will require international action to save the planet, but each country also faces the issue on the domestic front, within the context of its own geography and climate, political system, and socio-cultural norms and values. Canada shares much in common with the USA in being a wealthy nation sharing the same continent, but at the same time is distinct in terms of its economy, culture and politics. On the issue of climate change,

AAPOR 70th Annual Conference

Canada makes for a valuable case study as a country with a small population with one of the highest greenhouse gas emissions per capita on the planet. Climate change policy in Canada has taken a unique trajectory, starting with being among the first signatories of the Kyoto Accord, and later shifting to rear guard resistance that now consistently earns it “Dinosaur of the Year” awards from the international environmental community. This paper will present a concise history of Canadian public opinion as it relates to climate change, based on national surveys conducted by the author from the 1980s to the present (a new survey is being conducted in October 2014). The research reveals trends in opinions on such key issues as belief in the reality of climate change, concerns about climate change relative to other environmental issues, expectations about responsibility for actions, and support for specific policies, including carbon taxes (e.g., British Columbia introduced a true carbon tax in 2008). The research provides an instructive example of how public opinion can diverge significantly from government policy in a Parliamentary democracy.

Saving Energy: The Vital Role of Survey Research in Evaluating Low-Income Energy Efficiency Programs

Daniel Bausch, *APPRISE*

Kevin McGrath, *APPRISE*

Throughout the country, various states and utility companies have implemented energy efficiency programs designed to reduce home energy usage for low-income households. These programs seek to achieve sustained energy savings for qualified households by installing home weatherization measures, replacing inefficient energy-using equipment, or providing education to household members about energy usage and energy reduction opportunities. While these programs often share common features, they differ widely in their specific objectives, design features, implementation approaches, and outcomes. A primary challenge for low-income energy efficiency program administrators and evaluators is collecting needed data to effectively characterize and analyze a program’s performance. Data from program staff, home audits and inspections, and utility bills are important sources of information, yet these data do not provide key information on client perceptions, knowledge, and experience. This presentation will use data from multiple surveys completed with a wide variety of program participants to demonstrate how survey data helps to identify quantitative program impacts and provide valuable qualitative information to inform low-income energy efficiency program analysis. Specifically, we will examine how survey research assists with understanding two important areas. First, we will discuss how survey data provides information from participants about household comfort, safety, and health, including providing data from before and after program intervention. Second, we will review how survey data captures important information about participant perceptions of their energy usage, energy-saving behavior, and energy-related knowledge. We will present information on how this survey data can help compliment and inform the interpretation of other data, and we will discuss practical approaches to incorporating survey research in comprehensive program evaluations. This presentation will be a useful overview of low-income energy efficiency programs, and it will benefit researchers and policymakers seeking to understand how to incorporate survey research to better inform low-income program design and evaluation.

AAPOR 70th Annual Conference

The Effect of Question Wording on Measurement of Science Literacy

Aaron Maitland, *Westat*

Roger Tourangeau, *Westat*

Hanyu Sun, *Westat*

Yanna Yan, *University of Michigan*

Science literacy serves as a resource for individuals and communities. Since the late 1970s, the National Science Foundation has sponsored a series of surveys to gauge public attitudes toward and understanding of science and technology. The scientific community has raised concerns about some of the survey items that have been part of the factual knowledge scale for many years. There is growing concern that the NSF's factual knowledge scale has become less valid over time. For example, an item measuring knowledge of evolution ("Human beings, as we know them today, developed from earlier species of animals") increasingly seems to confound religious beliefs and scientific knowledge; a large number of fundamentalists are aware of the science, but nonetheless respond "false" or "don't know" to the item. This paper investigates the measurement properties of different versions of the evolution item. We examine two alternative wordings to the evolution item. The first alternative focuses the item more directly on the theory of evolution by prefacing it with "According to the theory of evolution..." The second alternative changes the focus of the item from the evolution of humans to the evolution of elephants. We use experimental data from the General Social Survey and two smaller methodological experiments to examine the effect of the alternative item wordings on the percent answering the item correctly and the effect on correlations with other items in the factual knowledge scale. Our findings show that both alternative item wordings increase the proportion of correct responses. In addition, confirmatory factor analyses demonstrate improved construct validity since the responses to the alternative item wordings are more strongly correlated with science knowledge and more weakly correlated with religious beliefs than responses to the originally worded item.

Panel: Quality of Qualitative Research: Setting Standards for Qualitative Public Opinion Research and Pretesting

Panel: Quality of Qualitative Research: Setting Standards for Qualitative Public Opinion Research and Pretesting

S. Andrew Stavisky, *Government Accountability Office*

While public opinion researchers are re-examining how standards are set and under what circumstances they need to be updated, we turn our focus to qualitative public opinion research. Though not a new area of study, qualitative research has long lived in the shadow of quantitative research in the field of public opinion. While AAPOR has well-established standards for quantitative studies, no standards have been widely accepted for qualitative studies. This panel brings together speakers from across disciplines who have worked on qualitative research and standards. They will discuss their own standards and how those standards could be applied within public opinion research considering both qualitative evaluative studies and qualitative pretesting.

Ensuring Quality in Qualitative Research Through Content Analysis: The GAO Approach

S. Andrew Stavisky, *Government Accountability Office*

Given its responsibility as the watchdog for the U.S. government, The U.S. Government Accountability Office (GAO) must provide Congress with timely information that is objective, fact-based, nonpartisan, non-ideological, fair, and balanced. Because any primary data we collect and analyze to support our conclusions and recommendations must be above reproach, we build rigorous quality standards into our audit processes. In addition to the quantitative methods that we employ on our projects, many of the studies we produce each year include qualitative research components, such as case studies, in-depth individual interviews, group discussions, focus groups, expert panels, and forums. Our highest-risk policy studies include additional layers of rigor to ensure the utmost quality of qualitative data. For example, when conducting comprehensive content analyses, we typically use two independent coders to code items rather than a passive approach of one person coding the data and another person verifying. Using two independent coders provides for the calculation of inter-rater reliability as an additional assurance that codes are valid and reliable. When there is not agreement between the two analysts, an independent third person is used to arbitrate any disagreements. Regardless of the type of software used, (N'Vivo, Microsoft Excel, Microsoft Access or other software programs), these additional steps provide a more "objective, fact-based, nonpartisan, non-ideological, fair, and balanced" methodological approach to the analysis. This paper will describe the various measures GAO takes to ensure quality to the qualitative research process through enhanced content analysis. We will use examples from our "gold standard" content analysis procedures and describe how these analyses are used in the reports we present to Congress.

Collaborative Approaches to Qualitative Reliability and Validity: Examples from Evaluation and Policy Research

Diane Purvin, *Yale University School of Medicine/The Consultation Center*

Quality control is critical to the legitimacy of qualitative studies in evaluation and policy research. Some traditional approaches to reliability and validity, such as triangulation, "member checks," and assessments of inter-coder reliability, may not be feasible in applied studies with logistical limitations. I present two examples of qualitative quality control, adapting traditional approaches via collaborations among project staff and participants. Example 1: Validity in Evaluation of Youth Experiences with State Social Services To assess how well a state social services program was meeting the needs of high-risk youth, the evaluation team conducted a series of focus groups with participants. Five groups were completed with a total of 28 participants, aged 13-23. "Member checks," in which study participants review codes and conclusions reached by the data analysts, were not possible due to IRB requirements of confidentiality. As an alternative, an additional group of similarly-situated youth (n=7) was recruited to conduct coding of the focus group data, in collaboration with the evaluation team. Example 2: Reliability in Analysis of Open-Ended Survey Data A global organization of experts in economic and public policy conducts an annual membership survey (n>1500) to help focus their work. A team of 3 analysts—a lead and two assistants—develops a coding framework, codes responses and analyzes themes for the open-ended questions. Their tight reporting deadline, combined with voluminous data, makes formal checks of inter-rater coding infeasible. Instead, initial codes for each question are developed by the lead analyst based on the first 100 responses. The questions are then divided between the two assistants, who complete the coding, adding and modifying as needed. The lead analyst reviews and revises before submitting the coded data to

AAPOR 70th Annual Conference

the organization's project team. Analytic themes are similarly developed and revised iteratively. The client organization's content-area experts also contribute and review.

Using Conjoint Analysis to Improve the Validity of Focus Group Results

Rebecca Quarles, *QSA Research*

Socially desirable responses are a major threat to the validity of focus groups. In light of this threat, many moderators include independent individual measures as part of their study design. Our firm has been using Conjoint Analysis and other Trade Off Analysis methods to help defuse this threat for studies that explore decision-making processes for the past 20 years. Topics have ranged from consumer decisions about home purchases to public policy on disposal of nuclear waste. Conjoint Analysis is particularly effective because it places the individual respondent in a realistic decision-making situation in which he or she must trade off valued features against each other, as well as cost, which can be monetary or non-monetary. Conjoint analysis yields an objective, quantitative measure of each feature/cost for each respondent (utility scores). When used in qualitative research, respondents complete the conjoint exercise either via laptop computers or, much less expensively, via a simple card sort exercise. We have developed a quick and easy scoring method that allows respondents to self-score the card sort results. These results are the springboard for discussion in the group. By using Conjoint Analysis as an individual exercise at the beginning of the group discussion, we have bypassed any discussion of importance of features and costs, which are often subject to the threat of social desirability. The Conjoint results are used as a conversational icebreaker. The discussion is structured so that respondents talk about the reasons behind their scores. During analysis, we compute utility scores for each feature for each respondent and discuss respondents' comments in the context of those scores. We will demonstrate administration and analysis of the results as part of the panel discussion.

The NSF Interdisciplinary Standards for Systemic Qualitative Research: Their Relevance Ten Years Later

Ronald Chenail, *Abraham S. Fischler School of Education Nova Southeastern University*

In Lamont and Whites' 2008 report, *Interdisciplinary Standards for Systematic Qualitative Research*, they shared findings from the 2005 National Science Foundation (NSF) funded workshop for representatives from NSF's Cultural Anthropology, Law and Social Science, Political Science, and Sociology programs. In the monograph they presented what they saw as strengths, standards, and opportunities for qualitative research then and in the near future. They organized their findings in four sections: (1) Qualitative Research Design and Methods; (2) Standards for Qualitative Research across Disciplines (in Anthropology, Law and Social Science, Political Science and Sociology); (3) Recommendations for Producing Top Notch Qualitative Research; and (4) Promising New Research Areas and Topics. As part of the panel, I will review the major recommendations of the report focusing on (a) their strengths of qualitative research, (b) their shared standards of qualitative and quantitative methods, (c) their standards unique to qualitative research, and (d) their "shared" standards for designing and evaluating qualitative research across the four disciplines; note some of the criticisms levied against the standards as being a "conservative challenge" to some forms of qualitative research; discuss the prevalence and impact of these standards ten years later; and suggest their relevance in the realm of contemporary survey and public opinion research.

International Development of a Quality Framework in the Qualitative Context

Karen Kellard, *The Social Research Centre*

Researchers have long debated the notion of quality in qualitative research and evaluation – how it is defined and measured – and have demonstrated a shared commitment to rigour, robustness and relevance but without a solid agreement on what this means in a qualitative context, and indeed to what extent such standards or measures can be formalised given the nature of qualitative enquiry. Some have rejected entirely the notion of criteria, some accept the need for broad guiding principles, and yet others subscribe to criteria parallel to those used in quantitative contexts. The author plans to work collaboratively with colleagues here in Australia, in the UK and in the U.S. to explore these issues and ultimately to develop a quality framework in the qualitative context. Drawing particularly on progress made in the UK and the U.S. over the last decade on this subject (for example, Spencer et al, Symon and Cassell, Meyrick, Lincoln and Guba, Roller and Lavrakas), the framework will cover: · four central principles of qualitative research (that it adds new knowledge or evidence, that the chosen research design is defensible, that the conduct is rigorous and that the findings are credible and clearly linked to the data) · quality at different stages of the qualitative research process (design, sampling, recruitment, fieldwork, analysis and reporting) – what kind of indicators or measures could we be looking at, how and why? How does this differ in face-to-face, telephone and online qualitative settings? · quality elements that run throughout the research process (relating to, for example, research ethics, neutrality, reflexivity, transparency, objectivity, skills and so forth). It is anticipated that the resultant guiding principles can be used by research practitioners, research commissioners and research users to provide a framework for the design, conduct and reporting of high quality qualitative research.

Web Panels: Recruitment and Retention

The Implications of Survey Experience and Panel Conditioning on Data Quality

Michael Henderson, *Louisiana State University*

D. Sunshine Hillygus, *Duke University*

Among the issues of debate and concern about the use of online panels for survey research is the possible cumulative effect of participation in many surveys over time. Does survey experience impact data quality? In many ways, the concerns parallels that of panel conditioning in longitudinal studies; yet, in the case of online panels, responses may be impacted by past survey experience even when the current survey has no longitudinal design. In this paper we attempt to disentangle length of tenure in an online panel from panel conditioning associated with participation in a longitudinal study. We make use of two election panel surveys administered at the same time, by the same vendor, and which include many of the same questions – the 2007-2008 Associated Press Yahoo New Election Panel Study (APYN) and the American National Election Study 2008-2009 Panel (ANES). A key difference between these projects is that while the APYN drew upon the traditional online panel (with variation in panel tenure), the ANES required recruiting a sample made entirely of new participants. Together these two projects allow comparison of a sizable number of fresh online panel participants to respondents with far more experience. Because both projects have longitudinal designs, the impact of experience with opt-in panels can be separated from panel conditioning. Initial analyses suggest that survey experience from either source has similar consequences for measurement error but appears to have contrasting implications for estimates of bias in the measurement of political attitudes and behaviors.

The Role of Device Type and Respondent Characteristics in Internet Panel Survey Breakoff

Allan L. McCutcheon, *Gallup Research Center*

Survey breakoff is less likely in online panels than in one-time web surveys – due, at least in part, to early panel attrition by less-committed panel respondents (see e.g., McCutcheon, Rao, and Kaminska, 2014). Breakoff in online panels, however, remains problematic, and may serve as a prelude to panel attrition. Interest in breakoff involving internet survey respondents has been accelerated by the relatively recent availability of paradata collection methods for web surveys (Peytchev 2009). In addition to respondent and survey design characteristics, it is now relatively easy to obtain data such as the amount of time taken per survey item (response latency), number of response changes, time of day that survey breakoff occurs, and other factors that may contribute to survey breakoff. Internet panels offer the additional advantage of accumulated information regarding respondent characteristics. This study examines data from 18 waves of the Internet component of the Gallup Panel, a multi-mode, probability panel of American households. In addition to demographic characteristics and survey design factors (e.g., question complexity, topic, number of questions, grid patterns, survey length), the analysis will include self-reports on Internet usage and computer sophistication, paradata and, especially, device type to explore factors related to survey breakoff. Preliminary analysis indicates that while long-term panel members are less likely to breakoff, past behavior and respondent-reported computer usage are important predictors – those who have broken off in the past appear, and those who report less computer usage, appear to be more likely to breakoff again. The study will consider the potential use of predictive models for survey breakoff in developing possible adaptive design (Groves and Heeringa 2006) interventions for Internet surveys that may prove useful in delaying survey breakoff.

Impact of Images on Survey Participation, Respondents and Online Panel Recruitment

Mingnan Liu, *SurveyMonkey*

Noble Kuriakose, *SurveyMonkey*

Jon Cohen, *SurveyMonkey*

Sarah Cho, *SurveyMonkey*

This presentation investigates how images shown to respondents before they enter a self-administered survey impact whether and how they respond. Using the SurveyMonkey “endpage”, a page that is displayed nearly three-million times daily to survey respondents who have completed a user-created survey, this presentation will examine if the type of image that is displayed on the endpage has a large impact on survey participation and whether different images appeal to different types of respondents. The endpage also serves as a recruitment source for SurveyMonkey Audience, a non-probability based internet panel. In addition to survey participation, using our vast dataset of millions of daily views, we will also discuss how images impact panel recruitment, particularly among those groups who are hard to reach.

Conversation, Rapport and Interaction: Effects of Interviewers and Respondents on Data Quality

The Impact of Rapport on Data Quality in CAPI and Video-Mediated Interviews: Disclosure of Sensitive Information and Item Nonresponse

Hanyu Sun, *Westat*

Rapport is generally described as a sense of connection, mutual comfort and ease of conversational coordination during an interaction (Foucault, 2010). Although there is no universally accepted way to measure rapport, the general consensus is that it is good for survey interviews and may affect the quality of the responses obtained (e.g., Foucault, 2010; Cassell & Miller, 2007). Technological advances in recent years have made video-mediated interviews more feasible and affordable; however, little attention has been paid to videoconferencing as a potential mode of data collection in survey interviews. In video-mediated interviews, the interviewer and the respondent can see and talk to each other via a video window. Although rapport-related verbal behaviors have been found to increase the disclosure of moderately sensitive information in face-to-face interactions (e.g., Dijkstra, 1987; van der Zouwen et al., 1991), it is unknown if rapport can be established to the same extent in video-mediated interviews, leading to similar levels of disclosure. I examined the impact of rapport on data quality in CAPI and video-mediated interviews with a laboratory experiment that varies the level of rapport, the mode of data collection, and the version of the questionnaire. Eight professional interviewers and 128 respondents participated. I found that (1) there was no significant difference in rapport ratings between video-mediated and CAPI interviews; that (2) the impact of rapport on disclosure depends on question sensitivity: when questions are moderately or less sensitive, rapport does not seem to affect disclosure; whereas when questions are highly sensitive, rapport leads to more rather than less honest responses; and that (3) the effects of rapport on item nonresponse depends on question sensitivity: when questions are slightly sensitive, rapport motivates respondents to invest more effort to respond, whereas when questions are moderately or more sensitive, rapport leads to more item nonresponse.

Interviewer Voice Characteristics and Data Quality

Nuttirudee Charoenruk, *Survey Research and Methodology program, University of Nebraska-Lincoln*

Telephone surveys frequently contain socially desirable, socially undesirable, and complex questions that tend to produce problems for respondents (Tourangeau & Yan 2007). Previous research has examined how interviewers in telephone surveys affect data quality of these types of questions, primarily through examining interviewer behaviors (Dykema et al. 1997). Questions remain, however, about whether an interviewer's voice characteristics also affect data quality. This study examines whether voice characteristics of telephone interviewers (measured by Praat software) are associated with data quality in socially desirable, socially undesirable, and complex questions. Data for this study come from the Work and Leisure Today Survey conducted by AbtSRBI (NSF SES-1132015), a RDD CATI survey. I used multi-level models to examine how interviewers' pitch, intonation, speech rate, and fillers affect data quality in twelve socially desirable, socially undesirable, complex, and neutral questions (three in each of the categories). Data quality indicators included item nonresponse, interviewer-related variance, and the intraclass correlation coefficient for all questions; the directional hypothesis of "more is better" for socially undesirable questions and "less is better" for socially desirable questions; and rounding for complex and neutral questions. I found interviewer's pitch and rate

AAPOR 70th Annual Conference

of speaking affected data quality in complex questions and questions that prone to socially desirable bias, but did not affect data quality in neutral questions. However, the effects varied by interviewer gender. Male interviewers were more likely to obtain better data quality when they read a question with higher pitched voices and with higher rate of speaking. In contrast, female interviewers who spoke with lower pitched voices were more likely to collect better data quality than those who spoke with higher pitched voices. Findings from this study have implications for selecting and training interviewers for CATI surveys and selecting voices for audio computer-assisted self-interviews with the goal of maximizing data quality.

Using Data Mining to Examine Interviewer-Respondent Interactions in Calendar Interviews

Robert F. Belli, *University of Nebraska*

L. Dee Miller, *University of Nebraska*

Leen-Kiat Soh, *University of Nebraska*

Tarek Al Baghal, *University of Essex*

Determining which verbal behaviors of interviewers and respondents are dependent on one another is a complex problem which can be facilitated via data mining approaches. Data are derived from the interviews of 153 respondents of the Panel Study of Income Dynamics (PSID) who were interviewed about their life course histories. Behavioral sequences of interviewer-respondent interactions that were most predictive of respondents spontaneously using retrieval strategies in their generation of answers were examined. We also examined which behavioral sequences were predictive of retrospective reporting data quality as shown by correspondence between calendar responses with responses collected in prior waves of the PSID. Attention was focused on four respondent retrieval strategies: parallel behaviors in which respondents spontaneously use a contemporaneous event from their past to cue an event in a different theme, timing behaviors in which respondents report when a past event happened, duration behaviors in which respondents report how long an event occurred, and sequential behaviors in which respondents report what happened earlier or later in time. Separately, the prevalence of these behaviors is associated with better data quality. Verbal behaviors of immediately preceding interviewer and respondent turns of speech were assessed in terms of their co-occurrence with each respondent retrieval strategy using a decision tree data mining algorithm called C4.5. Different interviewer-respondent interactional sequences were derived for each retrieval strategy, and each sequence was assessed in being associated with data quality. Analyses discover that better data quality more of a function of respondent engagement in using retrieval strategies rather than interviewers using retrieval probes. Interviewers' use of parallel probes is associated with poorer data quality, whereas interviewers' use of timing and duration probes, especially in tandem, is associated with better data quality. Data mining techniques are valuable to examine which interviewer-respondent interactions will benefit data quality.

AAPOR 70th Annual Conference

Interviewer-Respondent Interactions in Conversational and Standardized Interviewing: Results from a National Face-to-Face Survey in Germany

Felicitas Mittereder, *Michigan Program in Survey Methodology, University of Michigan-Ann Arbor*

Jen Durow, *Michigan Program in Survey Methodology, University of Michigan-Ann Arbor*

Brady T. West, *Institute for Social Research, University of Michigan-Ann Arbor*

Frauke Kreuter, *Joint Program in Survey Methodology-University of Maryland*

Frederick G. Conrad, *Institute for Social Research, University of Michigan-Ann Arbor*

In recent years, researchers have given greater attention to the interviewing techniques used in survey data collection. The field is presently dominated by one main technique, often referred to as standardized interviewing. This is a technique whereby an interviewer reads a question exactly as authored and reacts to respondent confusion by employing neutral probes, without giving the respondent any further information. Conversely, in the “conversational” interviewing technique, interviewers are given greater liberty to provide clarification in response to respondent confusion. Proponents of standardized interviewing argue that all respondents should be exposed to the same stimuli. Contrariwise, conversational interviewing may elicit higher data quality by making sure that each respondent understands the questions in the way they were intended. In this study, we explore respondents’ reactions to both interviewing techniques in a national face-to-face survey conducted in Germany, where interviewers were randomly assigned to use one of the two techniques for their interviews. The survey items underscore a range of question difficulty, from simple and well-defined to ambiguous and cognitively taxing. The interviews were all audio-recorded, and interactions were coded and tabulated into four respective groups of variables: questions or incidences of confusion, definitions read by the interviewer in response to confusion, definitions read without evidence of confusion, and neutral probes used in response to confusion. We find that respondents show more evidence of confusion when they recognize the interviewer’s ability to respond. We also explore the interesting case when a difficult question is met with ease by respondents assigned to a standardized interviewer, yet with confusion by those in the conversational group. We address another concern in conversational interviewing that interviewers help too much and provide definitions without any evidence of respondent’s confusion. We conclude with suggestions for practice and future research in this area.

Maximizing Telephone and Cell Phone Survey Participation

Inclusion of Cell Phone Households in the 2014 FDA Health and Diet Survey

Alan Roshwalb, *Ipsos Public Affairs*

Chung-Tung Lin, *U.S. Food and Drug Administration*

Agn  le S. Lawson, *Ipsos Public Affairs*

The Health and Diet Survey (HDS) is one of the FDA surveys to track and gather information on consumer awareness, attitudes and practices related to health and diet issues. The HDS has been conducted since the early 1980’s. The last wave of the study was conducted in 2008, but it did not explicitly include cell phone users and cell phone-only households. The 2014 HDS implemented an overlapping dual frame design to include cell phone households. This paper reports on the study design, the composition of the weights for the study, and the effects of design and its weights on some preliminary results. The paper will examine differences between

AAPOR 70th Annual Conference

household types for landline-only, dual service households and cell phone-only households. We will discuss decision choices for the weights and their effects on the weights and results. A secondary analysis will look at the effects of including cell phone households on selected HDS results.

I Think I Know You: Matching Local Area Codes in National Telephone Surveys

Heather Knappen, *Metrix Matrix Inc.*

Chris Horn, *Metrix Matrix Inc.*

This paper will present follow-up research from a 2014 AAPOR poster presentation with a year-long analysis on the effect of area code matching for telephone surveys in the commercial sector. The prevalence of caller ID technology used for personal call screening often impedes survey response rates among respondents who are unlikely to answer calls from unfamiliar telephone numbers. This paper looks specifically at whether area code matching helps overcome this issue by increasing response rates, limiting the overall number of call attempts to exhaust a contact record and resulting in operational cost reductions. In 2012, Metrix Matrix instituted a method for smart dialing within its in-house call center. Smart dialing relies on purchasing phone numbers within targeted area codes. These numbers, with the desired area codes, are displayed on outgoing calls made to the same, or geographically adjacent, area codes. The data presented in this paper will span a period of four years, from 2011-2014, in an attempt to capture the effect on contact and cooperation rates before and after the implementation of smart dialing. The sample being analyzed will include cell phone and landline respondents across various industry sectors such as utilities, business, government and healthcare. This study is one of the first that tracks contact and cooperation rates as well as the effectiveness of area code matching for telephone surveys. As the AAPOR community continues to evaluate the role and effectiveness of telephone surveys and interviewing, the use of smart dialing may be a promising tool to encourage participation among potential respondents.

Impact of Pre-Notices on Response Rate in a National RDD Study in Norway

Sofia Pinero Kluch, *Gallup*

Robert Tortora, *Gallup*

Ken J. Kluch, *Gallup*

As public opinion researchers grapple with response rates globally, Gallup has recently tested one method to increase response rate using mailed pre-notices to a portion of a nationwide sample. Pre-notices have been extensively studied by Dillman and others, however this pre-notice study adds the use of an RDD sample in Norway, thus examining the use of pre-notices with an RDD samples where basic demographic information is available. Pre-notices were tested in Norway where CATI response rates are troublingly low (6% in spring of 2014). Of the total sample of approximately 16,000, 1000 individuals received a paper pre-notice via post. The pre-notice was addressed to the individual registered as the phone's owner. Two different letters were sent, one for landline and another for the mobile sample. For the mobile sample, the respondent answering the phone was selected to participate. For the landline sample, the most recent birthday method was used for selecting the household respondent. The landline letter stated that the member of the household with the most recent birthday would be selected to participate. The impact of the Norway pre-notices were examined by completes, response rates, refusal rates, and compared for the mobile and landline samples. Potential respondents receiving pre-notice letters were more likely to participate in the study (13% completes; 6% completes for non-recipients). The refusal rate was also lower, 45% versus 62%. There was no

AAPOR 70th Annual Conference

difference between mobile and landline samples in the effect of the pre-notice for completes. The pre-notices did not differentiate response rate for mobile and landline. Among those not receiving a pre-notice, the mobile sample (46%) was significantly more likely than landline (43%) to refuse. For those receiving a pre-notice there was no difference (mobile 38% landline 42%). The findings indicate substantial benefits of a pre-notice: doubling completes and reducing refusals by 27%.

The Effect of Varying Incentive Amounts on Physician Survey Response

HarmoniJoie Noel, *American Institutes for Research*

Grace Wang, *American Institutes for Research*

Akash Desai, *American Institutes for Research*

Alison Huang, *American Institutes for Research*

Physicians are notoriously difficult to survey given the number of demands on their time and because they are frequently surveyed. Response rates to physician surveys average about 10 percentage points lower than general population surveys (Cummings, Savitz, and Konrad 2001). Pre-paid monetary incentives are more effective than promised monetary incentives or non-monetary incentives in boosting response rates in general and for physicians (Dillman 1978; VanGeest, et al. 2007; Kellerman and Herold 2001). However, the amount of the incentive that is most effective at increasing response rates is unknown. There has been a wide range of pre-paid incentive amounts used in previous research, varying from \$1 to \$50; however, VanGeest, Johnson, and Welch's 2007 meta-analysis found minimal differences between these incentive amounts. Our sample includes about 4600 primary care physicians or their staff who are working in clinical care in the United States. We excluded residents and physicians working in hospitals or government facilities. We conducted a two stage survey design with a short mailed screener to determine eligibility for the telephone survey. Participants were randomly assigned to receive one of three unconditional incentive amounts (\$2, \$5, or \$10) at the screener stage; a control group did not get any incentive at the screener stage. This paper will 1) examine the effect of using a pre-paid incentive versus no incentive on physician response rates to a short mailed screener, 2) assess whether the effect of an incentive on response to a screener survey varies by the incentive amount, and 3) examine the cost effectiveness of offering larger incentives. Preliminary results show that physicians who received any incentive amount were more likely to complete the screener than those who received no incentive, and that larger incentives were associated with higher response rates. These findings can help inform future physician surveys.

Use of a Reimbursement to Increase the Proportion of Pay-As-You-Go Cellphone Respondents

Marcus E Berzofsky, *RTI International*

Bo Lu, *Ohio State University*

George L. Couzens, *RTI International*

Caroline Blanton, *RTI International*

Kimberly Peterson, *RTI International*

Jamie Ridenhour, *RTI International*

Tim Sahr, *Ohio State University*

Robert Ashmead, *Ohio State University*

Amy Ferketich, *Ohio State University*

Tom Duffy, *RTI International*

As more persons move to using cellphones for their sole or primary telephone source, survey methodologists need to adapt and increase the percentage of respondents that come from a cellphone frame. At the same time, in order to reduce cellphone costs, many cellphone users are moving away from the traditional contract-based service plan to more flexible plans that are based on a pre-paid amount of minutes or pay-as-you-go arrangements (e.g., Tracfone). In states such as Ohio, it is estimated that approximately 30% of cellphone users are on such an alternative plan. Because these alternative plans cost a person plan minutes for both incoming and outgoing calls, persons on these plans may be less likely to participate in a survey for fear of running out of monthly minutes which may result in extra costs for the respondent. Excluding persons on these alternative plan types may introduce bias in survey estimates intended to be representative of the entire population or some low socioeconomic subpopulation. This concern over potential bias by under covering alternative phone plan persons is especially true for the 2015 Ohio Medicaid Assessment Survey (OMAS) which is designed to identify the Medicaid eligible population in the state of Ohio. While OMAS has not traditionally provided an incentive or reimbursement for their survey, the 2015 pilot study, which included 300 cellphone respondents, experimented with a \$10 reimbursement. This paper presents the results of a split-sample experiment which tested if such a reimbursement increased participation overall or increased the proportion of respondents that are on alternative cellphone plans. Alternative plan respondents were identified through survey questions as well as identification flags provided by the sample vender.