



OPPORTUNITY

through

DIVERSITY



2010 CONFERENCE ABSTRACTS

AAPOR 65th Annual Conference
May 13-16, 2010

Chicago Marriott Downtown Magnificent Mile | Chicago, IL

WAPOR 63rd Annual Conference
May 11-13, 2010

The Avenue Hotel | Chicago, IL

WAPOR Concurrent Session A

Wednesday, May 12

1:45 p.m. - 3:15 p.m.

Elections around the World

Brazilian Electoral Scenario in 2010: Changes and Continuities in Post-Lula Brazilian Voting Behavior

Márcia Cavallari Nunes, IBOPE Inteligência Pesquisa e Consultoria Ltda
(marcia.cavallari@ibope.com.br); **João Francisco Resende, IBOPE Inteligência Pesquisa e Consultoria Ltda** (joao.resende@ibope.com.br); **Malu Giani, IBOPE Inteligência Pesquisa e Consultoria Ltda** (malu.giani@ibope.com.br)

The objective of this study is to analyze the Brazilian voting behavior after two uninterrupted decades of national elections. The 2010 elections can set up the first complete cycle of alternation of power in the Federal Executive Branch and for the first time since 1989, Luis Inácio Lula da Silva will not be running, he who has always been the main candidates in the last five elections for the Presidency. This aspect, combined with the undeniable popularity that Lula has built over his two terms as President, will make the presidential election of 2010 as one of the most competitive in recent history and opens the possibility for a discussion of political proposals which will influence the Brazilian economic, social and politic environment in the coming years.

In addition to such aspects, we will investigate how changes and trends in contemporary Brazilian society will reflect in political behavior and elections in the coming years (ongoing changes such as the aging of the population; increase in the population's level of education; strengthening the national economy and the expansion of domestic consumer market; expansion of social policies etc.).

We will do an attitudinal segmentation of Brazilian voters, based on accumulated knowledge and literature on voting behavior, to increase the ability to anticipate trends and forecasts of the presidential results. Will be developed a quali-quantitative study of Brazilian voters and a socio demographic and attitudinal segmentation, using statistical techniques, in order to examine the profiles of each group and the relationships between these profiles with certain patterns of voting behavior.

All these aspects make it even more compelling the understanding of electoral behavior of Brazilians nowadays. That understanding is a significant contribution that the polls can provide for communication and political marketing in 2010 and beyond, contributing to the strengthening of Brazilian democracy.

“War and Peace” and Elections in Sri Lanka

Colin John Irwin, Institute of Irish Studies, University of Liverpool (colin.irwin@liverpool.ac.uk);
Pradeep Peiris, Social Indicator, Centre for Policy Alternatives (pradeep@cpasocialindicator.org)

With the defeat of the Tamil Tigers (LTTE) in 2009, a period of Sinhala triumphalism and Presidential and General elections in 2010, coloured by ethnic politics, can a new constitution provide an 'Opportunity through Diversity' that will herald a new era of peace, stability and economic prosperity for Sri Lanka?

A series of three polls conducted before the end of the war and after both elections tested the acceptability of constitutional reforms developed by the All Party Representative Committee (APRC) during the turbulent months of the Government's victory and two hotly contested electoral campaigns. Through the lens of a religious and ethnic demographic analysis problems in the reform proposals are identified and attempts made to suggest remedial political action to address the concerns of specific communities and constituencies.

With the addition of a political party analysis the prospects for coalition formation within the multi-party political environment of Sri Lanka is also examined. As constitutional change in Sri Lanka requires a two-thirds majority in Parliament such coalitions and pacts will be essential for the passage of such reforms. With this point in mind detailed analysis are made of each parties key manifesto policies and their own and other constituents support for them to map out all the possibilities for building alliances and consensus around a common reform package.

Finally a political analysis is made of the impact this program of applied polling had on the political process and how such transparent disclosure, following the end of the war and through the elections, influenced key political players and the decisions they had to make. Were the polls simply observing and noting changing public attitudes or did this program of research help to shape and move both public opinion and the position and role of political elites towards a constructive political accommodation?

All (Anti Incumbency Index): New Evidences and Examples From Around the Globe to Understand the Hypothesis and Quantification

Yashwant Deshmukh, Cvoter Foundation (Yashwant@Teamcvoter.Com)

In this paper, we will examine the validity of All (Anti Incumbency Index) hypotheses with latest trends of FPTP electoral verdicts around the globe. The democratic electoral verdict is about checks and balances that the public sentiment keeps on the ruling parties. Sometimes this sentiment is confused with the re-election of a certain ruling combination, which is incorrect. For example in the last UK elections where Labor came back to power suggesting a pro-incumbency verdict. The truth is far from that.

The countries with a positive incumbency index (including the conflict-ridden and democratically transitional states like Iraq, Bangladesh, Nepal, Congo, Pakistan, Afghanistan, Sri Lanka) are all struggling democracies. Having socio-economic indicators on the negative side, the struggling countries are trying to develop in a self-correction mode. But to have Socio-economic indices on negative side and the anti-incumbency index on the positive is a sure shot recipe for democratic disaster.

Even a rich country like Malaysia with sound economic fundamentals had a positive All till recent times, indicating as per our hypotheses a troubled democracy. But a change in All over the last two elections is indicating to better future for opposition parties and democracy at large. The latest elections in Japan achieved this phenomenon, where finally the opposition party swept to power for the huge All built up over the years.

When comparing the anti-incumbency with various socio-economic indices, it emerges countries with a stable All or healthy democracies also have a significantly stronger development. On the other hand, countries having a pro-incumbency bent display a relatively lower rate of socio-economic development. Failed or troubled democracies with low anti-incumbency sentiments have, on an average, poor social and developmental indicators.

Explaining the 2010 British General Election

Mark Gill, MORI Caribbean (markgill@moricaribbean.com); Robert Worcester, MORI Caribbean (sirrobertworcester@moricaribbean.com)

6th May 2010 is the "odds on favourite" to be the date of the next UK General Election. This paper will present one of the first comprehensive analyses of the outcome of the election using MORI's model of electoral behaviour, the Political Triangle ®. This model proposes that election results in the UK can be explained by voters' appraisals of the party leaders, the image of the main political parties and of the policy positions of the parties on voters' key issues. Polls during the course of the election campaign, together with historical trend data, will allow the authors to explore the impact of these three dimensions on voting behaviour, particularly among the floating voters.

The paper will also outline the result of the General Election through a detailed swing analysis, which measures the level of switching between two parties and provides a clear analysis of which groups of people, or types of seats, moved more or less from the average change.

In addition, to analyzing voting behaviour the paper will present findings on the performance of the opinion polls.

The paper builds on research and analysis of the last three UK General Elections, published in three books titled "Explaining Labour's Landslide" (1997), "Explaining Labour's Second Landslide" (2001) and "Explaining Labour's Landslip" (2005). The analysis for this paper and discussions at the conference will contribute towards MORI's fourth book on explaining the 2010 General Election.

Media and Public Opinion

Challenges in Spiral of Silence Research

Thomas Roessing, Institut fuer Publizistik, University of Mainz (roessing@uni-mainz.de)

Thirty-five years of spiral of silence research (Noelle-Neumann 1974; 1984) yielded important advances in understanding public opinion and with regard to methodology of surveys and content analysis (Scheufele/Moy 2000; Roessing 2009). However, there are still challenges for spiral of silence researchers left, concerning the areas of epistemology, theory, and methodology. It is the aim of this paper, to discuss some of these challenges and to propose approaches for further research.

An important epistemological problem of spiral of silence research is the fact that social reality changes over time and by place. While the phenomenon of the spiral of silence is believed to exist in all cultures at all times, the concrete processes of public opinion occur in different manifestations under different circumstances. This makes it so difficult to find generally applicable methods and to make concluding statements about the nature of public opinion.

Central theoretical problems lie in the field of minorities whose members are willing to talk (seemingly contradicting Noelle-Neumann's theory). There is data (e. g. from Israel), which suggests that the concept of involvement should be integrated into the spiral of silence theory. Another theoretical problem is online communication. How is public opinion affected by changing communication habits in a world that experiences globalization as well as localization at the same time?

Methodological problems concern the measurement of willingness to speak out and of the perception of opinion climates. While there is much creativity among researchers to invent new tests for willingness to talk, there are still problems with the measurement of the perceived climate of opinion. This is especially the case where interval or ratio scale level data is needed for statistical analysis.

Coverage of News Events Over Time

Stefan Geiss, Johannes Gutenberg-Universität Mainz (stefan.geiss@uni-mainz.de)

News reporting is contingent on news events. The media's and the public's attention mostly focuses on problems if there are events which exemplify these underlying problems. There are several typologies of news events, but none of these deal with the patterns of coverage of news events. This is surprising, because the patterns of coverage influence the amount of attention the public devotes to an issue. And only then, nation-wide public opinion formation is likely. This paper is concerned with three goals: (1) Identifying variables describing the patterns of coverage of news events; (2) Identifying types of news events which are homogenous regarding their patterns of coverage; (3) Testing whether specific kinds of events (e.g. crimes) exhibit typical patterns of coverage.

A content analysis of over 48,000 news items was re-analyzed by the author. It surveyed the political reporting of four prestige newspapers continuously over the year 2006. Events were identified as peaks in issue time-series. Over 200 peaks were found and 182 were identified as definite news events.

The most important findings regarding the research questions were: (1) Events can be accurately described by using three indicators: (a) the duration of coverage, (b) the peak number of news items, (c) the peak's position. (2) Four types of news events were identified: (a) Events with long-lasting coverage, and a high and late peak; (b) Events with long-lasting coverage, and a low centered peak; (c) Events with short-term coverage, and a low and very late peak; (d) Events with short-term coverage, and a low centered peak. (3) Some types of events (e.g. crimes, terrorist attacks, accidents) were limited to short-term coverage, while other kinds of events led to both long-term and short-term coverage (political proposals, elections, international conflicts, wars).

Need for Orientation, Media Use, and Agenda Melding

David H. Weaver, School of Journalism, Indiana University (weaver@indiana.edu); Donald L. Shaw, School of Journalism, University of North Carolina (cardinal@email.unc.edu); Bartosz Wojdyski, School of Journalism, University of North Carolina (bartw@email.unc.edu)

A "need for orientation" has been studied as a motive for media use, and voters with more of this need generally pay more attention to mediated political information. They also tend to reflect media agendas more than do voters with less need for orientation. But voters are not passive. We hypothesize that voters pick and choose among media topics (objects) and details (attributes), and meld them into a pattern that fits with their established political values and expectations. We also predict that voters with a higher need for orientation are more likely than those with a lower need to pay attention to a broader range of news media, and to construct issue and attribute agendas that are more reflective of these media agendas. This research aims to understand voters' media selection, media use, and the role of media in shaping their perceptions of presidential candidates and issues during the 2008 U.S. presidential campaign. The study centers on Chapel Hill, North Carolina, the location of the seminal 1968 media agenda setting study (McCombs & Shaw, 1972). In-depth interviews were conducted with 71 Chapel Hill voters in the months leading up to the 2008 election, and an extensive content analysis was done of six newspapers, 13 television programs, eight radio programs, nine web sites, and two magazines. This paper examines the relationships between participants' need for orientation, use of news media, and the perceived utility of various media platforms and outlets in learning about the issues and candidates of the campaign.

Role of War Blogs During the Lebanese-Israeli War in 2006

Dr. Ibrahim Saleh, The Arab European Media Observatory (isaleh@aucegypt.edu)

Every day, millions of bloggers share their opinions with a global audience discussing literally everything. What began as a hobby has changed the landscape for journalists and policymakers alike.

The current phase of globalization is marked with a mix of technology and war (War blogs) that has reshaped and redirected news media coverage in recent times in a big way. The earlier example was the first Iraq war in August 1990 and then the second Iraq war in 2003, and later the Israeli Invasion of Lebanon in 2006. Hence, war blogs ended the former camouflaged agony of people and the destruction of nation, by offering live coverage of the war that dumbed the viewers into awe.

War blogs are even more vibrant in the current agitations in the Middle East where those online journals featuring wry and candid observations about life in wartime, have become a serious political influence tool. For salient topics in global affairs, the blog-sphere functions as a rare combination of distributed expertise, real-time collective response to breaking news, and public opinion barometer.

During the Israeli invasion of Lebanon in 2006, war blogs have grown into full-fledged spheres of public deliberation and political commentary. Political blogs have blossomed into parallel feeds of sometimes ideologically aligned news and information. They are obvious web sites for examining nascent media frames about issues, that can feed back into the mainstream media to steer public debates.

Having said so, the research investigates the role of war blogs during the Israeli invasion of Lebanon in 2006. It aims at answering the questions of the public dependence on war blogs for news. The research analyzes the content and the discourse of the war blogs written in English that were dedicated to the issue under study, by examining their comprehensiveness as well as their news angles,

Methodology: Internet Surveys

How to Improve Response Rates to Web Surveys: Practical Guidance Based on Theory and Experimentation

Don A. Dillman, Washington State University (dillman@wsu.edu); Morgan Millar, Washington State University (morgan_millar@wsu.edu); Benjamin L. Messer, Washington State University (aluvial@gmail.com)

Conventional wisdom suggests that public opinion researchers must accept as fact that response rates to all forms of surveying are in decline. In the case of web surveys the problem is less one of decline than the fact that they have never been high, except for certain specialized populations, and there is little evidence that they are likely to become higher. This sense of reality has led to a shift in concern by many internet surveyors towards simply measuring whether non-response error exists and away from putting emphasis on getting higher response rates. In this paper we propose that the appropriate application of existing theories can provide guidance for improving web survey response rates significantly beyond current expectations. Doing so may help researchers in the reduction of sampling error as well as improving the credibility of surveys and the methods used. It may also help in the reduction of nonresponse error.

Several theoretical explanations of why people do and do not respond to various survey modes exist, including cognitive dissonance, social exchange, reciprocal obligation and leverage-salience. In this paper we draw from existing theories and research to propose means by which web response rates can be improved significantly over most current expectations. We do this by using theory-based arguments to propose methods for collecting survey information over the Internet, and provide evidence to support those arguments from experiments by the authors as well as others. We conclude with specific recommendations for improving response to web surveys and for advancing the testing of the theoretically-based propositions presented here.

Kids' Life and Times: Using the Internet to Give Children a Voice

Katrina L Lloyd, Queen's University Belfast (k.lloyd@qub.ac.uk)

Over the years, researchers from different disciplines have utilized a wide variety of research methods to give children a voice on issues that affect them. Particularly common are qualitative methods such as focus groups and small group discussions. Much rarer are large-scale quantitative surveys, which are a valuable way of comparing data from across different age-groups, countries and over time. Particularly lacking are surveys that give children an opportunity to voice their opinions using the technologies that play a major part in their lives such as the internet. To test the feasibility of collecting data from children using a large-scale quantitative survey on the internet, a pilot Kids' Life and Times (KLT) was carried out in June 2008 with a view to running the survey annually thereafter. All Primary 7 (P7) children (aged 10 and 11 years) in Northern Ireland were invited to participate in KLT and the data were collected using an animated questionnaire delivered via the internet in schools. Following consultations with P7 children in focus groups, the pilot survey included topics that were of importance to this age group such as bullying, health and education. The paper will discuss how well this method for collecting data from children worked in practice, the quality of information received and the response rate achieved. It will also consider lessons learned for future KLT surveys and for other surveys with children using a similar methodology. The presentation will include a live demonstration of the online animated questionnaire.

Non-Internet Households Who Were Given Internet Access to Participate in an Online Panel: Where Do They Go From There?

Annette C Scherpenzeel, CentERdata, Tilburg University (a.c.scherpenzeel@uvt.nl); Corrie Vis, CentERdata (C.M.Vis@uvt.nl); Klaas De Vos, CentERdata (k.devos@uvt.nl); Jorna Leenheer, CentERdata (J.Leenheer@uvt.nl)

A newly established Internet panel in the Netherlands, the LISS panel, is based on a probability sample from the population register, covering the general population. This sample thus includes households that have no Internet connection or computer. The panel provides a computer and internet connection to those households to participate in the monthly online questionnaires. In this paper we investigate the initial and the long-term representativity of these respondents for the total non-internet population. On the basis of a combination of register data and contact data we compare non-internet households that accepted the offer of a cost-free computer and internet access with the non-internet households who refused it. Furthermore, we investigate how much and in what way the group who took the offer starts to use the internet after been given the equipment and how that compares to households who had internet access before they were invited to participate in the panel. Finally, we look at the response behavior and attrition rates of this special group of panel members over time.

The Influence of Response Sets on the Results and Quality of Multivariate Analyses

Olaf Jandura, Ludwig-Maximilians-Universität München (jandura@ifkw.lmu.de); Hans-Bernd Brosius, Ludwig-Maximilians-Universität München (brosius@ifkw.lmu.de)

In this research project we will demonstrate the vulnerability of several statistical tests and procedures towards response sets. Using data sets from two standardized surveys (face to face & web), secondary data analyses were conducted. Given the importance of the mode of data collection, as it can alleviate or exacerbate response sets, we used data gathered in face-to-face interviews as well as via mail and online surveys. As expected, we found that the overall number of response sets was low, with the highest number found in the online survey. Further testing showed, however, that despite their low number response sets are far from negligible: whilst simple bivariate tests such as cross tabulations and means response sets are hardly affected due to the low share of invalid answers results of multivariate procedures are considerably altered. Our results show that a share of 0.001 percent (or 24 of 19.800 respondents) of respondents using response sets change the results of cluster analyses (k-means)

significantly. Based on these findings, we finally discuss how to identify and deal with respondents who used response sets.

Multi-Country Surveys

Changes in Support for European Integration in Five Political Systems

Marijn van Klingeren, University of Amsterdam (ASCoR) (m.vanklingeren@uva.nl); Hajo G. Boomgaarden, University of Amsterdam (ASCoR) (h.boomgaarden@uva.nl); Claes H. de Vreese, University of Amsterdam (ASCoR) (c.h.devreese@uva.nl)

The project of European integration is moving beyond a mere economic endeavor to encompass more facets of political and social life. In line with this development, in recent literature on public support for European integration, explanatory variables are argued to have shifted from so-called 'hard' factors (such as economic expectations) to 'soft' factors (such as feelings of a national / territorial identity or immigration attitudes). Even though by today it appears that these 'soft' predictors are most relevant for explaining public support, it remains uncertain if they have in fact gained in importance over time and if the importance of 'hard' factors has weakened. The main interest of this study is therefore to empirically address this assumed shift.

Attitudes on anti-immigration are often used as a predictor of European integration support. In this study anti-migration attitudes are utilized as explanatory, as well as dependent variables. Believing that nationalistic feelings are at the base of European integration and immigration support, it is expected that these elements are explained by similar factors.

The study draws on exiting data sources (Eurobarometer) from five European political systems (Wallonia, Flanders, the Netherlands, Denmark and Sweden) between 1990-2010. The countries are similar in various ways; they all have a multi-party system, stable and peaceful governance, similar media systems, and similar social conditions (e.g. number of immigrants/ asylum seekers). Because of these macro level similarities it is possible to rule out certain macro explanations beforehand, and focus on the dynamics of the importance of different micro level indicators. The results of the study are discussed in the light of recent literature, and so a clear shift in importance from 'hard' factors to 'soft' factors in the five systems is expected, though at varying intensity given variation in the salience of immigration and identity related attitudes.

Differing Degrees? Public Opinion on Climate Change Across Nations

Sandy Marquart-Pyatt, Michigan State University (marqua41@msu.edu)

The issue of climate change is uniquely positioned at the nexus of the physical and social sciences. Its expansive audience includes scientists, policy-makers, government officials, citizens, and public opinion researchers, underscoring the importance of multiple perspectives for understanding how views about climate change are being created and sustained comparatively and cross-nationally. Important questions remain about the global reach of climate change and how to explain its influences, especially as reflected in public opinion data. This research chronicles public opinion on climate change across more than fifty countries using the World Values Survey. An investigation of primary influences on views of climate change focuses on the extent to which political polarization arguments from research on the United States translates across nations, in addition to socio-demographic and attitudinal indicators. Results demonstrate more similarities than differences across countries with regard to factors underpinning opinions about climate change.

From Theory to Practice: Towards a Means of Measuring World Opinion

Frank Louis Rusciano, Rider University (rusciano@rider.edu); Michael Brogan, Rider University (mbrogan@rider.edu)

This paper attempts to design a means of measuring world opinion, based upon two sets of theoretical assumptions. The first notes that common usages of world opinion in international newspapers all contain the same components when samples are combined; however, their constructions of the content and meaning of the concept vary according to interests and cultural interpretation when broken down by region or nation (Rusciano, 1998) . Hence, these variations in meaning must be considered when measuring world opinion. The second is articulated by Christopher Hill (1996), when he distinguishes between international public opinion as the opinion of nations (as usually reflected by their leaders' or their nations' "official positions" on an issue) versus the opinion of individuals (as reflected by international polls in individual countries) . Hence, both meanings must be considered when measuring world opinion.

To deal with these problems, we measure world opinion on the United States' image in the world using three different approaches applied to a Pew International Survey. The first aggregates all of the data from the 40 nation study together, and analyzes the factors that affect the US's image (for an example, see Wilcox, et. al., 1993). The second approach correlates mean values on various attitudes that may affect evaluations of the US with the mean evaluations themselves, using the nation as the unit of analysis (for an example, see Goldsmith, et. al., 2005), again analyzing the factors that affect the US's image. The third approach addresses Rusciano's problem by including the relationships among the correlated attitudes, using the nation again as the unit of analysis (for an example, see Rusciano, 2003) analyzing effects on the US's image. The conclusion suggests that a mixed approach best addresses these theoretical problems for describing the communication and flow of opinions on a given subject

Paranormal Beliefs in East Asia

Jibum Kim, NORC (kim-jibum@norc.org); Seokho Kim, Sungkyunkwan University (seokhok@skku.edu); Hongjoon Yoo, Sungkyunkwan University (yhjcti@hanmail.net)

Despite much interest in paranormal beliefs in the West, little is known about paranormal beliefs in the East. We examine the prevalence and correlates of paranormal belief in China, Japan, Korea, and Taiwan.

Using the 2008 East Asian Social Survey (EASS), we first assess the prevalence of two paranormal beliefs: using a fortune-teller for naming children and caring about auspicious or ominous days for special occasions such as weddings, moves, and funeral ceremonies. Then, using logistic regression, we analyzed the correlates of these paranormal beliefs.

Our preliminary result shows that the prevalence of using fortune-tellers for naming children is 17% in China, 51% in Japan, 42% in Korea, and 64% in Taiwan. Also, the prevalence of considering very much/fairly about auspicious or ominous days for special occasions is 67% in China, 75% in Japan, and 38% in Korea, and 74% in Taiwan. A consistent factor for paranormal beliefs across four countries is religious identification. With the exception of the Chinese in terms of using fortune-tellers for naming children, those who believe in Eastern religions, such as Buddhists, are more likely to believe in paranormal beliefs than non-religious people. But, Christians are even less likely than non-religious people to believe in paranormal beliefs.

We show that the prevalence of paranormal beliefs varies by the specific item in the East Asian countries. Unlike the findings in the West, religious identification in the EAST appears to be related to paranormal beliefs.

Status of Women in the Middle East and North Africa (SWMENA): A Grassroots Research and Advocacy Approach

Rola Abdul-Latif, IFES (rabdullatif@ifes.org); Lauren Serpe, IFES (lserpe@ifes.org)

In this poster, IFES researchers will present findings from comparative research in Lebanon, Morocco, and Yemen that strives to use survey research to capture key data on women's status in the economic, political, social, and health sectors in these countries. Women in the Middle East and North Africa (MENA) lag behind women in most of the rest of the world, and behind men in their region, in areas such as their social and economic autonomy, labor force participation, political representation and health. However, the paucity of accurate and comprehensive national data in these areas, especially data disaggregated by gender, poses a problem in adequately evaluating the status of women in these countries. It also hinders the ability of local NGOs and women's groups in effectively influencing legislators and other policy makers and advocating on behalf of women.

Given the general lack of national census-type data on these issues in these countries, the Status of Women in the Middle East and North Africa (SWMENA) project is utilizing a survey-based approach to fill the gaps in existing research and data on women.

This project has engaged local NGOs and researchers on women's issues in evaluating existing data on women, identifying the gaps, and helping to define new measurement concepts. National surveys have been subsequently designed and are in various stages of implementation to capture much-needed data and information on the status of women. The survey questionnaires comprise both a core section that is standard across the three countries and a distinct country module that dealt with country-specific gender issues.

The newly available survey data will be used to generate composite social indices that will summarize key themes and allow comparisons across regions and communities within the same country and between the three countries under study.

WAPOR Concurrent Session B

Wednesday, May 12

3:45 p.m. - 5:15 p.m.

Democratic Values

Albania 2009: Polling Challenges and Opportunities in Emerging Democracies

Zeljka Buturovic, Zogby International (zeljka@zogby.com); Katherine Schwalbe, Zogby International (katy@zogby.com); Grace Ren, Zogby International (grace@zogby.com)

In a run up to 2009 Albanian parliamentary elections, Zogby International, in collaboration with a local vendor, conducted five national surveys and the exit poll altogether covering 17,000 participants. We also made a prediction for the allocation of 140 parliamentary seats according to D'Hondt and Sainte-Lague methods mandated by Albanian law. The 2009 parliamentary election in Albania featured two main parties: the incumbent Democratic Party of Albania and the opposition Socialist Party of Albania as well as a number of minor parties, most of which eventually coalesced into coalitions around the two main parties. The polling in Albania presented a number of challenges, most of which were mitigated by a relatively high level of competence and integrity of the local vendor. Due to low telephone coverage, we had to rely on a random route sampling method and face-to-face interviewing. Our research was featured prominently in the local media, leading to pressure from minor political parties. The estimation of the error in the parliament seat assignment presented significant computational challenges and we were not able to do a proper bootstrap procedure. This problem was augmented by the fact that a separate seat allocation needed to be performed for each of the 12 counties. Our final national poll predicted a 45: 42 win for Democratic Party over the Socialists with the third party LSI getting 5% of the vote. The final results had Democrats at 47%, Socialists at 45%, and LSI at 6% of the vote. However, exit polling overestimated the margin of error for the win in terms of parliamentary seats, predicting an advantage of 14 seats for the ruling coalition, only 4 of which materialized. This was most likely due to unwillingness of opposition supporters in this young democracy to share their views with the exit poll interviewers.

Crime and Erosion of Democratic Values in Latin America: Evidence From the Americas Barometer 2008 Surveys

Orlando J. Perez, Central Michigan University (perez1oj@cmich.edu)

In the Latin American context of extremely high crime, political scientists and policy makers alike need to ask whether crime, and the associated fear of crime, is helping to build popular support for repressive measures, and by extension, for repressive regimes. In short, is it possible that growing crime is a threat to the durability of democracy in Latin America? To test this proposition, this paper uses the 2008 AmericasBarometer surveys in 21 nations of Latin America. The premise of the paper is that rising crime and insecurity undermine democratic values and increase support for authoritarian measures. As crime rates increase and governments fail to stem the tide, citizen's belief that democracy is the best system may decline. Some citizens may support the implementation of greater controls or extra-judicial measures. High levels of crime may reduce levels of tolerance and interpersonal trust, thus undermining social capital. Finally, crime victimization and the fear of crime could drive citizens to lose faith in their political institutions, particularly the police and judicial authorities.

Giving the Citizenry a Greater Voice in Government

Steven Kull, Program on International Policy Attitudes, U. of MD (skull@pipa.org)

International polling reveals high levels of dissatisfaction with governments in democratic states: dissatisfaction is even positively correlated with Freedom House ratings. Such dissatisfaction appears to be highly related to the perception that governments do not understand and are not responsive to the views of the citizenry. This paper will review international polling on this problem and possible solutions. It will assess options for giving the citizenry a greater voice in government, with special emphasis on the potential for establishing large scale Citizen Advisory Panels that would participate in in-depth surveys, interact online, and participate in citizen assemblies. To be discussed would be: how such panels could be designed, developed and financed, how they could interact with the government, and possible roles they could play in government.

Rule of Law or Rule by Optimism? The Case of the Andean Region

Ana Lucia Cordova Cazar, Gallup Research Center - University of Nebraska Lincoln (al.cordovacazar@huskers.unl.edu); Francisco Lopez Bermudez, ADA (lopezbermudezf@yahoo.com); Rene Bautista, UNL - Gallup Research Center (rene.bautista@huskers.unl.edu); Lei Huang, UNL - Gallup Research Center (leihuang03@hotmail.com); Allan L McCutcheon, UNL Gallup Research Center (amccutch@unlserve.unl.edu); Tarek Albaghal, UNL - Gallup Research Center (mtalbaghal@hotmail.com)

Since the 1980s, when most Latin American countries replaced their military-authoritarian regimes with elected governments, the idea that the rule of law is essential for national development has been a basic premise in politicians' and policymakers' discourse (Carothers, 2010).

Already in the 1960s, the "Law and Development Movement" had given a strong political backing to the idea. They firmly believed that law (reform) and judicial systems could contribute to the development of societies (Merryman, 1977). Likewise, scholars coming from different fields, such as institutional economists (Coase, 1960), and economic historians (North & Thomas, 1973) stressed the influence that property rights and contract enforcement have on development. Political scientists (Rodrik et al. 2002) have also viewed these institutions as essential in the development process.

Still, the formal acceptance of the idea of the rule of law has not been able to avoid painful economic and political experiences in Latin America. The Andean region is a case in point, with several financial crises, and a high proportion of ousted governments and "re-founded" countries.

Using data from the Gallup World Poll, a probability-based multi-national survey, the relation between public support for the rule of law and development is analyzed in the six Andean countries: Bolivia, Colombia, Chile, Ecuador, Peru and Venezuela.

Preliminary results, using questions specifically targeted to the region measuring respondents' perceptions about the way the rule of law works in their nations and future expectations about their countries' development, show a significant correlation between these two variables, although the degree of association varies across countries (Bolivia: 0.1658; Colombia: 0.1042; Chile: 0.1419; Ecuador: 0.1148; Peru: 0.1427, and Venezuela: 0.1932). Interestingly, the public's perceived degree of rule of law in each nation also varies in counter-intuitive ways.

Support for Free Expression in Mexico: A Nationwide Survey Compared With New U.S. Data

Robert Wyatt, Middle Tennessee State University (rwyatt@mtsu.edu); Jorge Rojas, Independent Political Consultant (miyorch@yahoo.com); Jason Reineke, Middle Tennessee State University (jreineke@mtsu.edu); Ken Blake, Middle Tennessee State University (kblake@mtsu.edu)

Support for Free Expression in Mexico: A Nationwide Survey Compared with New U.S. Data

Mexico, like all other representative governments, may be, in poet Mario Vargas Llosa's characterization, an "imperfect democracy." Great strides have, however, been made in assuring pluralistic representation of various parties and factions since the 2006 election brought to an end Llosa's "perfect dictatorship" dominated by the Partido Revolucionario Institucional throughout most of the 20th century.

In order to provide an assessment of Mexican citizens' support for speech and media rights, a nationwide survey was conducted in Mexico in spring 2009, after the elections. The survey sampled approximately 1,200 respondents selected randomly by household and interviewed face-to-face. This study represents, in part, a replication of a series of international studies of support for free expression first conducted for the American Society of Newspaper Editors to celebrate the 200th anniversary of the U.S. First Amendment in 1991 (Wyatt, 1991). This study has also been replicated in Israel among Arabs and Jews, in Moscow, and in Hong Kong (Andsager, Wyatt, & Martin, 2004).

Preliminary results based only on Federal District results indicate that Mexican support for free expression is similar to that in other democracies under stress, particularly Israel. Notable was rather low support for the individual right to speak out for any candidate for office one favors. If, as anticipated, citizens of the D.F. are more liberal in their support for democratic values than other Mexicans, our nationwide survey may reveal important challenges for democratization in Mexico.

Results from the Mexican study will be compared on key issues with a new survey of U.S. free expression attitudes in order to provide an ongoing portrait of support for democratic values in the fluid environment of Mexico's emerging democracy.

Media Effects

Cultivation and the Spiral of Silence: Theoretical and Empirical Intersections and Their Implications for Public Opinion Research

James Shanahan, Boston University (shanahan@bu.edu); Dietram Scheufele, University of Wisconsin – Madison (scheufele@wisc.edu)

This paper explores theoretical intersections between two important theories of media effect: cultivation theory and the spiral of silence. While both theories were put forth at about the same time (late 60s/early 70s), and both deal with the role that media play in establishing boundaries and guiding public discourse, they have rarely been considered as complementary or competing accounts of public opinion formation. Both theories were originally put forth as something of a return to stronger media effects, after a long period in which it was common to discount or downplay media effects. Surprisingly, despite these congruencies, the theories have rarely been used in tandem when explaining media effects, and only one study has probed possible intersections and complementarities (Shanahan, Scheufele, Yang & Hizi, 2004).

In this study, we use data from a national survey (N=1,040, RR3: 41%) to empirically test explanatory models based on cultivation and spiral of silence theorizing. In particular, we are interested in testing which theory provides a better account of public responses to emerging technologies, with a particular focus on nanotechnology.

Because nanotechnology is a complex scientific issue with moral implications to some of its applications, it receives wide media coverage (a necessary condition for cultivation effects) and is also an ideal issue for study within the spiral of silence model. Thus, both cultivation and spiral of silence are positioned to present theoretical narratives on how people respond to nanotechnology. Also, we explore whether the theories are complementary, including whether media use and other variables typically examined in cultivation theory establish moderating mediating conditions for spiral of silence processes.

Managing Media and Political Opportunities at Home and Abroad: A Case Study of Mexican Americans

Andrea Elizabeth Hickerson, Rochester Institute of Technology (aahgpt@rit.edu)

While periods of migration are woven throughout ancient and modern history, today new media technologies encourage transnational ties more than ever before. Media such as satellite television, cell phones and the Internet, facilitate interpersonal communication across national borders, make news from the “homeland” accessible worldwide, and allow for the creation and dissemination of media content exclusively for consumption by transnational communities. Communication research has devoted considerable energy to understanding how communication behaviors can affect civic and political participation, but has not considered how transnational practices may impact these relationships. Using two data sets of Mexican Americans, this study tests a series of hypotheses and research questions emerging from literature on communication and transnational studies. The first is a secondary analysis of a large national telephone survey of 5704 Latinos of Mexican descent; the second comprises of original data from a convenience sample of 323 people of Mexican descent in Yakima, Washington. Results suggest that communication behavior is a better predictor of political participation in the U.S. than in Mexico. Furthermore, with regards to communication behavior in general, respondents demonstrated a preference for local news, and there was some evidence that people who used media for information-seeking purposes and felt “more American” were more likely to be political active than those who felt “more Mexican.”

Social Networks, Community Interaction, and Civic Participation of United States Immigrants
Wenlin Liu, Department of Communication, University of Washington (wenlin08@uw.edu)

Though democratic theorists recognize the importance of an active citizenry, past research has largely overlooked the civic participation of immigrant populations—a nation’s new and future citizens. This research paper examines the factors that promote social and political incorporation of immigrant communities in the United States, focusing on the role of social context in the process of immigrants’ political re-socialization. Specifically, it proposes studying the interconnections between the properties of U.S. immigrants’ social networks, the ways in which they interact within their local communities, and their level of political knowledge and civic participation.

Based on a statistical analysis of data from the Current Population Survey: Voting and Civic Engagement Supplement 2008, my research hypothesizes that the nature of immigrants’ social networks and their corresponding social interactions within local communities will influence their level of civic engagement. Furthermore, I argue that the impact is realized through both an informational effect (increases in political

knowledge) and a mobilizing effect (increases in political campaign involvement and other types of civic activities), and the length of residence positively moderates the above causal relationships.

A hierarchical multiple regression and path analysis is conducted to delineate the influence path between two sets of social contextual variables (social networks and community interaction) and the level of political knowledge and immigrant civic participation. Besides, variables from the traditional SES model of political participation (e.g. education level, income) are re-assessed. In conclusion, its theoretical contribution to the model of immigrant political socialization is discussed.

The Differential Role of Personal Contact for Generalized Perceived Threats Towards Muslims and Immigrants

Anouk van Drunen, University of Amsterdam (a.s.vandrunen@uva.nl); Hajo Boomgaarden, University of Amsterdam (h.boomgaarden@uva.nl); Rens Vliegenthart, University of Amsterdam (r.vliegenthart@uva.nl); Claes H. De Vreese, University of Amsterdam (c.h.devreese@uva.nl)

In many places immigrant integration has become a continuous issue. In the Netherlands, the discussion of this topic is characterized by a shift from a multicultural perspective to one that emphasizes the need for assimilation of minority groups (Coenders et al, 2008). Information people receive from e.g. the media is an important factor of forming attitudes toward immigrants (Boomgaarden, 2007). Vasterman (2005) discusses how media can suddenly generate high levels of media attention on one specific topic and this seems true in the Netherlands with MP Geert Wilders and Muslim integration.

Information in the media can make certain interpretations or selection of cues more cognitively accessible. In the media negative frames often prevail and threats are often coupled with Muslims and immigrants. We study the effect of coupling the immigration topic with negative frames in the media on people's opinion. We do so with a 'decoupling experiment' (Sniderman et al, 2004), embedded in a general population survey (n=856) conducted in 2009, which measures the threat perception. Specifically we study the role of personal contacts. The effect of personal contacts depends on the nature of the contacts; it can help to identify with out-groups and as a result create more positive attitudes. In each condition of the experiment five statements were posed, representing a specific type of threat. Both the experimental conditions (reference too Muslims / immigrants as cause of threat) differ from the baseline condition (no reference). The mean 'perceived threat' is lower when a reference is added, which was counter to expectations. However, the distinction of type of threat is less important in the experimental conditions. There is a more general 'perceived threat' of immigrants and especially of Muslims. Personal contact has significant main effects and a few significant interaction effects. The direction depends on the type of contact.

Methodology: Selection, Weighting and Non-Response

Investigating the Errors that Occur with Within-Unit Respondent Selection

Paul J Lavrakas, Independent Consultant (pjlavrak@optonline.net); Trevor N Tompson, The Associated Press (ttompson@ap.org); Robert G Benford, GfK Custom Research NA (robert.benford@gfk.com); Christopher J Fleury, GfK Roper Public Affairs & Media (christopher.fleury@gfk.com)

Coverage of the general population in a household survey depends not only on the unit-level sampling design, but also on the method used to select an eligible respondent – the so-called “designated respondent” – within each household. In 2004, researchers from Westat (Rizzo, Brick, and Park)

described a new method for within-unit respondent selection based on the number of eligibles in a household, which in theory appears to be an accurate and cost-effective approach to respondent selection. Our paper will report on our latest use of a hybrid version of the Westat method that was implemented in more than 10 state and national RDD surveys by The Associated Press and GfK Roper in 2008, 2009, and will be implemented upcoming 2010 AP surveys. The method generates a much better overall unweighted female/male selection of eligibles than other non-Kish selection methods the AP has used. However, unexpected results consistently occur whereby there was a significant interaction between (a) gender of the interviewer and (b) gender of the respondents selected. The pattern observed in the 2009 samples was consistent with what we observed in the 2008 RDD landline surveys, and was found to be associated with various types of misselection errors and a form of respondent-related measurement error. For the 2010 samples we are building in the capture of more paradata to help us investigate the role of differential nonresponse in explaining the fact that female interviewers have a significantly more skewed respondent gender selection than do male interviewers.

Optimizing Samples for Media Consumption in Sub-Saharan Africa
Mark W. Andrews, Survey Consultant (mandrews@cox.net)

In theory, sample optimization is a fairly simple matter, but in practice it can be difficult or impossible. One difficulty is that, the theory relies on knowing the variance of a single item when most surveys target a wide range of measures. It is often necessary to develop metrics or other means to pool variance across a number of variables. A second difficulty is that the theory assumes knowledge about the variance of the variable of interest. In some cases, data from previous studies provides estimates of this variance, but in other cases a researcher can only guess using data from comparable countries.

It is the purpose of this paper to develop a strategy for optimizing the samples for media studies in Sub-Saharan Africa across three strata (capital, non-capital urban, rural). The paper will propose a method for pooling variance across key measures using multivariate analysis. The paper will then develop and compare optimized samples for a number of countries in Sub-Saharan Africa. Finally, the paper will examine the potential for applying optimization strategies derive in one country or a group of countries onto a similar country for which no sufficient media consumption data is available.

The Mahalanobis Distance as a Pre-Election Poll Accuracy Measure
Stefanos Dris, CERN and N.T.U. Athens (stefanos.dris@cern.ch)

In order to entirely capture the amount of pre-election poll error in elections with more than two candidate parties, it is necessary to use a metric that includes the prediction errors of the smaller parties that gain a significant percentage of the vote. This is especially relevant to countries that employ (at least partial) proportional representation, since the vote share of every candidate contributes to the final parliamentary seat distribution. Modified versions of Mosteller's third method from the Social Science Research Council's report on the 1948 US pre-election polls have often been used to address this issue. However, researchers have disagreed on the exact methodology, citing the obvious limitation that inclusion of many smaller parties will artificially reduce the amount of calculated error. In order to eliminate the need for an arbitrary decision on how many parties to include, this paper makes the argument for using the Mahalanobis Distance (MD), which is the multivariate extension of the univariate z-score. The polling process can be approximately modelled as the sampling of a multivariate normal distribution. A poll's statistical distance from the actual election result is then given by the MD. Its applicability as a poll accuracy measure is demonstrated, along with its relationship with the chi-square statistic. The methodology for calculating the MD is presented in the context of the 2009 Greek parliamentary election and its advantages and disadvantages discussed. Finally, possible extensions based on Shipman's

'interval measure' (2009) are considered, with the aim of preserving its mathematical objectiveness when assessing polls that include undecided voters.

Weighing the Options: Are There Substantial Benefits to Weighting to the Number of Potential Respondents in a Household?

Mark W. Andrews, Survey Consultant (mandrews@cox.net)

A common methodology for selecting respondents to participate in a survey is to select a household at random and to select an eligible respondent at random once contact is made with the household. If this methodology is employed, an individual from a household with many eligible respondents is much less likely to be selected than an individual who is the only eligible respondent. From a purest statistical perspective, this should be reflected in probability of selection (POS) weights. Many, however, purposefully or unwittingly skip this step.

This paper will examine the impact this practice has on the quality of data using a number of studies in Sub-Saharan Africa. First, the paper will discuss the types of biases that this practice may introduce given different cultural contexts. Second it will examine the performance of pure POS weights in a number of countries in terms of a) estimating baseline demographics b) weight magnitude. Finally, the paper will assess the value of POS weights as a preliminary step in a non-POS weighting strategy. In particular, it will assess if using pure POS weights as an initial step will improve or hinder the effectiveness of a complex raking strategy using pre-determined demographic targets.

Religion

Clashing Civilizations and the "Faith Factor": The Influence of Religion on Anti-American Sentiment in Western Europe in the Post-9/11 Period

Erin Carriere-Kretschmer, Pew Research Center (ecarriere-kretschmer@pewresearch.org); Jody C. Baumgartner, East Carolina University (jodyb@jodyb.net)

Recently published research found that religious beliefs play a significant role in predicting American public opinion on foreign policy issues in the Middle East (Baumgartner, Francia, & Morris 2008). In particular, the authors found that Evangelical Christians have remained strong supporters of a hawkish foreign policy toward the Middle East, even as overall public support for the Iraq War declines. In addition, Evangelicals are among the strongest supporters of Israel and hold more negative views of Islam than others. Does this "faith factor" hold for non-American publics as well?

Using data from the Pew Research Center Global Attitudes Project we examine the influence of religion and religiosity on views of the U.S. and US foreign policy toward the Middle East among British, French, German and Spanish citizens in the post-9/11 period. In particular, we focus on the attitudes of the Muslim minority populations within these countries, expecting to see significant differences in anti-American sentiment between this group and non-Muslims. In addition, we expect that the more faithful Christians in these countries are more likely to hold positive views of the U.S. and be more hawkish when considering Middle East policy.

A Measure of Religiosity Applicable to All Faiths

Jon D. Miller, Michigan State University (jdmler@msu.edu); Ronald Inglehart, University of Michigan (rfi@umich.edu)

Religion is an important factor in the lives of many people on this planet and it has become an important variable for social scientists. Traditionally, many surveys asked about religious affiliation, including more specific denomination within larger traditions. Some surveys have asked about the frequency of attendance and, more recently, more surveys are asking about the frequency of prayer. There is little agreement about the best measure of religiosity. The World Values Survey has included questions about religion in all of its five waves and the fifth wave in 2005-7 included numerous items that relate to various aspects of religiosity. The proposed paper will describe the results of a confirmatory factor analysis (using LISREL) for a set of seven items. The seven items were combined into a factor score-based index that ranges from zero to 10. The proposed paper will describe the distribution of this index in 50 countries, ranging from very religious to non-religious. The index will also be assessed for its fit in predominately Roman Catholic, Greek Orthodox, Muslim, Buddhist, and Protestant countries. We will look at the relationships between our Religiosity Index and selected political, social, and family behaviors and values across the 50 countries.

Religious Change Around the World

Tom W. Smith, NORC (smitht@norc.uchicago.edu)

Religious change around the world is a complex phenomenon. No simple description such as secularization, religious revival, or believing without belonging captures the complexity of the process. The pattern varies across countries, indicators, and periods.

This paper presents the most comprehensive analysis to date of global religious trends. It examines the follow types of data regarding religious beliefs and behaviors:

1. Time series data from
 - a. Adult and youth samples in the United States
 - b. Adult cross-national studies
2. Age-cohort differences from
 - a. The United States
 - b. Cross-national studies
3. Development and religious beliefs and behaviors
 - a. Country-level correlations between developmental and religious indicators
 - b. Scientists and religion

Religious Commitment and Extremist Political Views in Nigeria

Neha Sahgal, Pew Research Center (nsahgal@pewresearch.org); Kathleen Holzwart Sprehe, Pew Research Center (kholzwart@pewresearch.org)

Is religious commitment related to extremist political views? While the study of religious extremism has gained increased ground in recent years, relatively little is known about the social bases of the phenomenon. Studying religious extremism poses several challenges for scholars: the utility of terms such as "moderate" and "extremist" have been keenly debated, further, lack of reliable data on these issues creates measurement problems. Using survey data, we study the impact of high levels of religious services attendance along with other demographic variables such as sex, age, education as well as level of satisfaction with the way things are in the country on support for several different political views among Muslims in Nigeria. Specifically, we study support for: implementing Shariah law, allowing judges to use their religious beliefs when deciding family and property disputes, punishments like whippings and cutting off of hands for crimes like theft and robbery and stoning people who commit adultery. Given the history of religious conflict in the country, application of Shariah law in northern areas, and current events surrounding the "Nigerian Taliban," and the government, Nigeria is a particularly important case to study

extremist views. Using logistic regression models, we find that religious service attendance does not have a significant impact on support for the above mentioned policies. Regardless of religious commitment or education, region is found to have a significant impact on support for these policies. In fact, in the case for support of Shariah law, Muslims in the northern region, where aspects of Shariah law are already in place are nearly twice as likely to support its implementation than Muslims in the south.

The "Sufi-Sunnis": Measuring Islamic Affiliation in Sub-Saharan Africa

Allison Pond, The Pew Forum on Religion & Public Life (apond@pewforum.org)

While some consider Sufism to be a genuine sect of Islam, others have labeled Sufis “non-Muslims,” leading to their persecution in several Muslim countries. Sufism has a tenuous history with Sunni Islam, but a new survey finds that in several sub-Saharan African countries, a high proportion of self-identified Sunnis indicate that they also belong to a Sufi order.

Using data from a new Pew Research Center survey of religion in nineteen countries in sub-Saharan Africa, we examine the social, political and religious views of these “Sufi-Sunnis” in Nigeria, Chad and Djibouti. We compare them to Sunnis who do not identify as Sufis, looking specifically at their religious beliefs and practices, their attitudes towards religious pluralism and extremism, and their views of democracy, theocracy and religious freedom. Our results underscore the unique characteristics of Islam on the African sub-continent and illustrate that Islam, far from being a monolithic religion, is quite internally diverse. These findings also raise important questions about how we measure Muslim affiliation and the implications for future research on Islam, both in Africa and on a global scale.

WAPOR Concurrent Session C

Thursday, May 13

8:15 a.m. - 9:45 a.m.

Methodology: Survey Response

Branching Out: A Comparison of Branching Response Formats with Single Response Formats

Randall K. Thomas, ICF International (randall.k.thomas@gmail.com); William Bryan Higgins, ICF International (bhiggins@icfi.com)

Unlike single response formats that use a single dimension to measure attitudes, the branching technique of attitude measurement separates the rating task into two different sequential tasks for measuring bipolar attitudes: direction of attitude (positive versus negative) and intensity of attitude (strength). Some research (Malhotra, Krosnick, and Thomas, 2009, POQ) has indicated that superior outcomes are obtained with branching that has more articulation (3 categories rather than 2 categories) on the endpoints rather than in the middle. However, no comparisons were made with single response formats. Further, the nature of the midpoint may have affected the results. We report here 2 studies comparing a number of branching alternatives, including a consideration of alternative middle responses, with a series of single response measures to determine relative efficacy of the scales. In Experiment 1, we had respondents who indicated their liking and approval for President George W. Bush in the U.S. as well as the extent to which they thought that spending for the military should be increased or decreased using 1 of 5 response formats (2 branching with 2 or 3 endpoints and 3 different single response formats). In Experiment 2, we extended the study internationally in 6 countries (US, UK, France, Germany, Spain, Italy) and used a different set of topics – attitudes and behaviors toward a number of different actions. Overall, we replicated Malhotra et al. in Experiment 1, but also found that single response format took less time and was generally similar in criterion-related validity. In addition, we found that different transformations of the criteria were associated with differential validity of the response formats.

Cultural Values and Survey Response Style

Timothy Johnson, Survey Research Laboratory (timj@uic.edu); Allyson Holbrook, University Of Illinois At Chicago Survey Research Laboratory (allyson@uic.edu); Sharon Shavitt, University Of Illinois At Urbana-Champaign (shavitt@illinois.edu); Young Ik Cho, University Of Illinois At Chicago Survey Research Laboratory (youngcho@uic.edu); Noel Chavez, University Of Illinois At Chicago (nchavez@uic.edu); Saul Weiner, University Of Illinois At Chicago (sweiner@uic.edu); Kelly Mazarno, University Of Illinois At Chicago Survey Research Laboratory (kellym@srl.uic.edu)

A small body of research now suggests that respondent cultural orientations may be associated with systematic variation in response style behaviors. Much of this work has relied on aggregate, national level measures of cultural values, rather than individual differences measures, which may be more sensitive to individual respondent variability in propensity to employ response styles when answering survey questions. In this paper, we measure sets of cultural values identified within the Hofstede, Schwartz, and Triandis conceptual frameworks to investigate cultural effects on three response styles: extreme response style, middling response style and acquiescence. A diverse, stratified sample of 400 African American, Korean American, Mexican American and non-Hispanic white adult respondents interviewed in Chicago in 2008-09 are used to examine two sets of hypotheses. The first set of hypotheses directly relates

cultural values with response behaviors. A second set investigates the degree to which these cultural values mediate the well-documented relationships between race/ethnicity and each of these response styles. All hypotheses are tested using structural equation models that construct latent measures of both response style behaviors and cultural values. Implications for the conduct of multi-ethnic surveys and recommendations for future research are discussed.

Freelancers vs. Employees: Interviewer Employment Type and Data Quality on the Spanish Survey of Household Finances

Sarah M Hughes, NORC (hughes-sarah@norc.org)

In Europe, survey interviewers are often hired as temporary workers and paid by the completed case. In some instances, survey organizations hire and supervise the interviewers directly, in other instances survey organizations subcontract work to groups of supervisors and interviewers, providing the necessary materials, sample, and training but leaving the day-to-day case management to the subcontractors. On the third round of the Spanish Survey of Household Finances, conducted in 2008-2009, one group of interviewers was supervised directly by the survey organization, while the majority of interviewers were supervised by subcontractors. This employment structure provides an opportunity to explore whether employment type has an effect on data quality and production. Using specific data quality measures included in the response data as well as process data gathered during the fieldwork, this paper demonstrates the effect of direct supervision of interviewers on data quality and, to some extent, production. Some comparative data on interviewer employment type is drawn from countries participating in the European Social Survey, and comparative data on quality measures is drawn from the US Survey of Consumer Finances.

Pseudo-Opinions in Arts-Affairs: Prevalence and Social Determinants in Mail and Questionnaire Surveys

Karl-Heinz Reuband, University of Duesseldorf _ Soziologie II (reuband@phil-fak.uni-duesseldorf.de)

Pseudo-opinions have been studied in the past on the basis of face-to-face surveys and mostly with regard to questions dealing with political issues (Bishop 2005). Social desirability effects have usually been seen as a driving force for giving a "substantive" response to questions on fictitious issues. Whether anonymous surveys, which are less affected by social desirability effects than face-to-face interviews, also produce pseudo-opinions has less often been studied. In the paper the issue will be dealt on the basis of two studies, both based on random samples: a local population survey that was done by a mail survey (N=1044) and a study of museum visitors which was done by a questionnaire survey (The museum questionnaire could be deposited either in the museum or sent back to the researcher) (N=986). The fictitious issue concerns the knowledge and rating of a non-existing artist.

Public Opinion Around the World

A Survey of Iranians: Attitudes Toward the U.S. and Internal Political Issues

Sara Beth Elson, RAND Corporation (belson@rand.org); Alireza Nader, RAND Corporation (anader@rand.org)

Tensions between the United States and Iran over a wide range of Iranian activities- including the alleged development of nuclear weapons and the ensuing threat to Israel as well as the continuing repression of Iranian citizens- makes the understanding of Iranian public opinion, popular pressures on the regime, the

domestic political environment in which the regime operates, and perceptions of the United States, more important than ever before. As such, we are conducting a telephone survey across all provinces of Iran to examine the attitudes of a relatively affluent cross-section of Iranian society toward the United States, toward the idea of developing nuclear weapons, and toward the Iranian economy, among other topics. The survey also examines which issues were important to affluent Iranians as they voted in the June, 2009 presidential election. Finally, the survey includes multiple open-ended questions designed to elicit affluent Iranians' thoughts on re-establishing relations with the United States, the history of Iranian-U.S. relations, and other topics. Ultimately, the survey is designed to improve our understanding of the political climate inside Iran and how relatively affluent individuals inside the country are willing or able to speak about issues of concern to their society.

Gender, Policy, and Public Opinion in Syria and Lebanon

John D Willingham, D3 systems (john.willingham@d3systems.com)

Lebanon and Syria share a border, as well as a relatively heterogeneous ethnic and religious societal make-up. They also share a contentious history with one other. In the wake of French colonization, each state embarked on radically different paths to becoming a modern nation-state. At times, women were the focus of public debates over the future of each country.

The status of women in Syria and Lebanon has long been considered crucial to public discourses regarding the condition of each country. This paper provides insight into how women perceive their de facto status in the face of competing discourses shaping the political, economic, and social landscape of each country. Namely, it looks at the behaviors and attitudes of Syrian and Lebanese women today, while also examining efforts made by each country's government to protect and promote women's rights.

As part of a collective effort to understand the attitudes and behaviors of Muslim women around the world, D3 Systems, Inc. sponsors and manages the Women in Muslim Countries (WIMC) study. WIMC is designed to measure women's empowerment in actual daily practice, providing an in-depth look into the oft-perceived gap between current public policy and empowerment initiatives, as well as actual practice on the personal and local level. The answers are intended to yield a metric for promoting excellence in public policy by informing policymakers on women's attitudes about the affects of policy initiatives, as well as a measure of the degree of effectiveness of those policies.

The Syrian and Lebanese surveys consist of interviews with randomly selected Muslim women. The respondents were interviewed in person by native speakers. Syria had n=505 interviewed in person May 13-24, 2008; Lebanon n=508 interviewed in person October 15-20, 2008. Each study has a margin of error of +/- 4% at the 95% confidence level.

Public Evaluation on the Leadership Performance of the United States and China in East Asia

Lei Huang, SRAM, University of Nebraska-Lincoln (leihuang03@hotmail.com); Rene Bautista, SRAM, University of Nebraska-Lincoln (rene.bautista@huskers.unl.edu); Allan McCutcheon, SRAM, University of Nebraska-Lincoln (amccutch@unlserve.unl.edu); AnaLucia Cordova Cazar, SRAM, University of Nebraska-Lincoln (analucordova@yahoo.com); Tarek Albaghal, SRAM, University of Nebraska Lincoln (mtalbaghal@hotmail.com)

China, the "strategic competitor" of the United States, is peacefully increasing its influence worldwide, especially in East Asia under its "win-win strategy". Although the political and military relations among countries in East-Asia were unstable historically, China's rising economic power establishes a close tie with countries in this region, enabling it to share power with the United State since the Asia Financial

Crisis, and becoming an economic powerhouse in Asia. Experts in international relation have been in a heated dispute regarding China's leadership status and its potential threat to security and stability in Asia. The present research assesses public opinion in four East-Asian nations/regions toward China and the U.S. using data from the Gallup World Poll. The World Poll includes a number of core questions about public support for the leadership performance of the two nations, as well as public political preference. In this research, public opinion in Hong Kong, Japan, South Korea and Taiwan are analyzed to describe the characteristics of the supporters of China and the US, using socio-demographic attributes and respondents' political preferences. Preliminary analysis indicates that the US still gains more support than China from Japan (30% vs. 13%), while China acquires more support from Hong Kong (64% vs. 22%) and South Korea (45% vs. 40%). However, in Japan, Taiwan and South Korea, people with higher levels of education are more likely to favor China's leadership performance (e.g., in Japan, those with the highest education level are more likely than those in the lowest level to support China, 24% vs. 8%).

The Impact of Military Operations on Afghan Optimism

Karl G Feld, D3 Systems, Inc. (Karl.Feld@d3systems.com); Anne Pessala, D3 Systems (anne.pessala@d3systems.com); John Richardson, D3 Systems (john.richardson@d3systems.com); Brianna Augenreich, D3 Systems (brianna.augenreich@d3systems.com)

D3 Systems has conducted nationwide, face-to-face public opinion surveys in Afghanistan every month since September 2005. Not surprisingly, public optimism about living conditions in the present and the future declines as military operations sweep through Afghans' homes and villages. However, as military operations cease, public optimism recovers to its previous levels. Opinion also only declines in those localities where operations are occurring, while other parts of the country aren't impacted.

This paper is an analysis of Afghan public optimism about the present and future as it has been affected by ISAF military operations. Analysis of provinces and regions is provided, associated with military operations of matching geographic scope. D3's monthly data collection beginning in late 2005 is used to track the consistency of the rebound phenomenon in Afghanistan.

Afghanistan's media environment, standard of living and the nature of Taliban and ISAF strategic communications may be causative themes for this phenomenon, among others. They are explored in the paper. Conclusions are drawn regarding short-term and long-term impact of ISAF kinetic and non-kinetic military operations on winning and maintaining the support of the Afghan people.

D3's Afghan nationwide surveys consist of interviews with randomly selected Afghan adults aged 18+ throughout Afghanistan. Respondents were interviewed in person by native Afghans. Interviewers were gender-matched with respondents. D3's survey is n=2000 a month with an associated margin of error of +/- 2% at the 95% confidence level.

Survey Research: Past, Present and Future

Does Familiarity Breed Contempt? Measuring and Comparing Survey Attitude Among New and Repeat Respondents Cross-Culturally

Edith Desiree de Leeuw, Utrecht University (e.d.deleeuw@uu.nl)

Declining response rates worldwide have stimulated interest in understanding what may be influencing the decline, and how it may vary across countries and survey populations. In this paper we develop and

test a short nine-item survey attitude scale that measures three constructs, thought by many to be related to decisions to participate in surveys, i.e. survey enjoyment, survey value, and survey burden. This survey attitude scale is based on earlier work by multiple authors (Cialdini (1991), Goyder (1986), Singer (1998), Stocke (2006), and Rogelberg et al (2001).).

The survey attitude scale was implemented in four quality online panels, two in The Netherlands and two in Germany, enabling us to compare survey attitude across cultures.

One of the Dutch and one of the German panels are recently established, whereas the other two (comparison) panels have been in existence for years. This provides unique data to investigate whether being asked to participate repeatedly in surveys, an obligation connected to all of the panels, is related to how respondents view their participation in surveys. In addition we present some data about the relation between survey attitude and nonresponse occurring in the panels over time.

Our overarching goal with this study is to develop an effective measure of how individuals are responding to surveys that might provide insight into steps researchers can take in order to improve survey effectiveness.

Stability and Reliability in the General Social Survey Panel, 2006-2008

Tom W. Smith, NORC (smitht@norc.uchicago.edu)

The General Social Survey (GSS) is in transition from a strict, replicating cross-section design to a design using rolling, short-term, three-wave panels. From 1972 through 2006 each GSS was based on a new sample. Under the new design there will be three components to each biennial GSS, a new starting panel, the first reinterview of the panel from the previous GSS, and the second (and last) reinterview with the panel from the next earlier GSS. Thus, in 2010 the target is to interview a) 2,000 respondents in the new panel (who will then be reinterviewed in 2010 and 2012), b) as many of the 2,023 2008 GSS respondents as possible (who will then be reinterviewed again in 2012), and c) as many of the 1,536 respondents from the 2006 GSS who were previously reinterviewed in 2008.

The analysis will look at the 2,000 GSS respondents from 2006 and consider a) panel attrition – how the panel respondents differ from the full sample and b) among those doing the GSS reinterviews in 2008, the stability and reliability of over 500 items measured in the GSS. Drawing upon test/retest studies carried out by the GSS in the 1970s, one focus will be on how the level differs by the type of variable (e.g. demographic, behavior, attitude, etc.), the format of the items (e.g. number of response categories), and the topic or substance of the items (e.g. politics, well-being, health, etc.). A second focus will be how the level of change differs by people's attributes (e.g. whether stability is greater among the young than the old).

Time to Go Beyond Demography in Reporting Public Opinion

Hans Zetterberg, ValueScope AB (hlzetterberg@hotmail.com)

In one crucial respect, polls remain at the stage where national opinion surveys started in the 1930s. Pollsters then took a clue from another nationwide enterprise, the census. The categories of age, sex, and (in the United States) race became what latitude and longitude are to geographers: stable references into which researcher place new discoveries. Modern theories of state, legitimacy, and democracy have little or no use for public opinion polls reported in demographic categories. Students of “participatory democracy” need reporting of opinions by degree of participation in different social movements, not age, sex, and other demographics. Students of “deliberative democracy” need to know opinions in interrelated networks. Students of legitimacy need reports of opinions by makers, keepers, brokers, and receivers of

cardinal values such as knowledge, wealth, order, sacredness, virtue. Students of development and contention need opinions classified by societal realms. This paper specifies such needs and hints at methodologies that can meet them.

Trust and Society

A Comparison of Women's Perceptions of Their Rights in Common Practice in Northern Sudan and Egypt

Eric Chappell, D3 Systems (Eric.Chappell@d3systems.com); Brittany Shanks, D3 Systems (Brittany.Shanks@d3systems.com)

The northern regions of Sudan are particularly linked to Egypt. Egypt's influence has been felt in the country since the Pharaohs and has been especially prominent since the creation of modern Sudan. While Egypt and much of Northern Sudan share an ideological past, Muslim women, an often overlooked segment of the population, express very different opinions regarding their quality of life, financial situations, and future outlook. This poster highlights key differences not only in the attitudes of Muslim women, but also how demographic statistics also play a role in highlighting similarities and differences between the two populations.

The Women in Muslim Countries (WIMC) Survey, sponsored by D3 Systems, strives to provide a voice for females residing in Muslim nations and to communicate their opinions and needs to public policymakers and the world. Data has been collected from national surveys completed in Sudan, Egypt and 21 other countries.

The general perception of the Sudan is of a nation struggling to form its identity amidst civil unrest and outside interference. As the poster demonstrates, female Sudanese have a significantly bleaker outlook of the future and their own abilities to influence it than Egyptians. This is in agreement with current demographic data that describes a population with a lower standard of living and lower income. Similarly, employment opportunities are low and are likely to remain so for these women. With these factors viewed as parts of the whole, the Sudanese female's outlook can be considered significantly less inspiring than her Egyptian equivalent.

The respondents of both surveys were randomly-selected Muslim women interviewed in person by native speakers. Sudan had n=698 interviewed from July 29-August 18, 2009; Egypt n=500 interviewed March 10-April 1, 2007. Each study has a margin of error of +/- 4% at 95% Confidence Level.

How Trust Influences Environmental Activism Across Nations. **Sandy Marquart-Pyatt, Michigan State University (marqua41@msu.edu)**

The role of trust in a variety of social settings and interactions has received a good deal of attention in recent years, and has been linked with a number of different behaviors. For instance, trust is widely believed to underlie citizen support for government and participation in various aspects of the political process. Yet its influences and potential consequences are not well understood, however, at different scales and directed toward specific topics, institutions and organizations. This is especially the case with regard to environmental issues, conditions, and concerns. In this research, therefore, factors influencing trust linked with environmental issues and conditions are investigated, highlighting how this type of trust is related to pro-environmental attitudes and behaviors. Specific attention is paid to the role of context in shaping the levels, sources, and consequences of trust in the attitude-behavior nexus. This research

examines trust in sources of environmental information as mechanisms through which individuals engage in environmentally related political actions in nineteen countries, emphasizing a comparison of advanced industrialized and former state socialist countries. Investigating the dynamics linked with trust cross-nationally extends prior research, adding an additional dimension to thinking about the global importance of environmental issues and the importance of context. While some similarities can be observed, notable differences also exist. Results are discussed in light of future cross-national studies with particular attention to how the findings might intersect with policy creation and implementation.

Methodological Problems with Surveying Trust

Susanne Lundasen, Ersta Sköndal University College (susanne.lundasen@gmail.com)

Social trust has been described as a central variable for explaining a variety of societal phenomena (Uslaner, 2002, Zak & Knack, 2001). The aim of this paper is to address some central problems related to surveying generalized trust or social trust. Many authors have pointed at the difficulties that may arise from the interpretation of the survey question on general trust from Noelle-Naumann (adopted by GSS, WVS and EVS to mention a few) (Milner & Mitamura, 2001; Zmerli, Newton & Montero, 2007). Problems that are mentioned in the literature are for instance problems of which heuristics the respondents use (are most people people you know or strangers?) when answering the general trust question; problems associated with the scales (cautious is not an opposite of trust) (Yamagishi, 1998); problems of context sensitivity of the question (Smith, 1997). Van Deth has also argued that the theoretical heterogeneity of the research on social capital/trust is not found in empirical research. This paper will however take the reasoning one step further through evaluating different measurements of social trust coming from a unique survey containing different types of trust questions. Questions of self reported trusting behaviour, assessments of trust towards different group and the generalized trust question are compared and analyzed. A second issue that is raised is if we even can survey distrusters as they often because of general lack of trust refuse to participate in surveys?

Trust in Justice and the Legitimacy of Legal Authorities: Pilot Data From a Major Study of European Public Opinion

Jonathan Jackson, London School of Economics (j.p.jackson@lse.ac.uk); Mike Hough, Kings College London (mike.hough@kcl.ac.uk); Stephen Farrall, University of Sheffield (S.Farrall@sheffield.ac.uk)

Public trust in their justice systems, public compliance with the law, and public cooperation with legal authorities – these are crucial to the maintenance of social order, as well as important indicators of a society's health and well-being. A major study is underway to examine legitimacy and trust in the police and courts across Europe. A European Commission FP7 funded project entitled JUSTIS: Scientific Indicators of Confidence in Justice, and a connected 50-item rotating module in Round 5 of the European Social Survey (ESS), will soon collect survey and contextual data from 30 or more European countries. In preparation for this, the current paper reports on the findings from JUSTIS pilot work (cognitive interviewing in five countries, and representative-sample surveys of four countries) and ESS pilot work (representative-sample surveys of two countries). The pilot data are discussed in the light of a conceptual and theoretical framework that sketches out the meaning and significance of public trust and institutional legitimacy in diverse contexts across Europe.

Trust in Scientists: The Role of Media in Establishing Trust in Sources of Information About Nanotechnology

Ashley A Anderson, University of Wisconsin-Madison (aaanderson3@wisc.edu); Dietram A Scheufele, University of Wisconsin-Madison (scheufele@gmail.com); Dominique Brossard, University of Wisconsin-Madison (dbrossard@wisc.edu)

The technological acceleration of modern society in conjunction with the rise of complexity in information exemplifies the paradox of democratic societies. Individuals face the potential risks and benefits of swiftly advancing technological innovations at the same time that sources of information about these technologies are multiplying and difficult to follow. In the face of this complexity, individuals turn to institutional trust to form attitudes and make judgments about emerging and controversial technologies (Brossard & Nisbet, 2007; Liu & Priest, 2009). The connection members of the public have with institutions and their corresponding representatives is often through media sources.

In this context, this study explores the relationship of media usage and trust in different institutions as purveyors of information on nanotechnology. Using public opinion data collected in the United States in 2007, the study specifically explores the potential link between public affairs media use and science media use and trust in scientists and in government. The study's results reveal that individuals who pay more attention to science and public affairs media are more trusting of scientists, but neither public affairs media use nor science media use is significantly related to trust in government. Deference to scientific authority is an important predictor of trust in both types of institutions—government and scientists—as sources of information about nanotechnology. Possible explanations and implications for public opinion about controversial technical innovations are discussed.

WAPOR Concurrent Session D

Thursday, May 13

10:15 a.m. - 11:45 a.m.

Challenges in Survey Methodology

Assessing Asian-Language Translated Questions About Self-Reported Racial/Ethnic Discrimination in a Health Survey

Martha Stapleton Kudela, Westat (marthakudela@westat.com); Kerry Levin, Westat (kerrylevin@westat.com); Jocelyn Newsome, Westat (jocelynnewsome@westat.com); Salma Shariff-Marco, National Cancer Institute (shariffs@mail.nih.gov); Nancy Breen, National Cancer Institute (breenn@mail.nih.gov); Gordon Willis, National Cancer Institute (willisg@mail.nih.gov)

Recent research has increasingly examined relationships between psychological and physical health, including the impact of discrimination experiences on the physical health of minority groups (Landrine, Klonoff, Corral, Fernandez, and Roesch, 2006; Paradies, 2006; Kressin, Raymond and Manze, 2008; Williams and Mohammed, 2008). Early pre-testing of discrimination items is an important step in assessing comprehension and other response processes, especially when the items have been translated into languages other than English. This paper seeks to extend existing research on the impact of cultural communication patterns on the survey responses of Asian respondents (Ma, 2003; Grandy, 1996) by using behavior coding methodologies to identify survey response problems these respondents have when answering questions about their discrimination experiences.

The National Cancer Institute is conducting a second field test of two versions of a Discrimination Module (DM) in the California Health Interview Survey (CHIS). One version asks about racial/ethnic discrimination specifically, attributing unfair treatment to race within each survey question. The other asks about unfair treatment first and follows up later with questions about the cause of the unfair treatment. For behavior coding purposes, we are audio-recording some of the Korean-, Vietnamese- and Chinese-language (Cantonese and Mandarin) interviews. We are also recording a small subset of interviews that are being conducted in English with respondents who have self-identified as being a member of one of these sub-groups. (Total n=400)

We will discuss how behavior coding results from this and a previous field test helped to determine which DM version to use in future administrations of the CHIS. We will discuss differences in survey interaction behaviors among the Asian-language sub-groups, including their reactions to being asked about their discrimination experiences. We will also discuss the implications of our findings for future surveys that explicitly include Asian sub-groups, and future applications of cross-cultural behavior coding.

Implementing Real-Time Surveys in Developing Countries:

Lessons From a Cross-National Health Opinion Survey in Five Developing Countries

Kana Fuse, Institute for Health Metrics and Evaluation (kananan@gmail.com); Ali Mokdad, Institute for Health Metrics and Evaluation (mokdaa@uw.edu); Emmanuela Gakidou, Institute for Health Metrics and Evaluation (gakidou@uw.edu); Andrew Dyer, ICF Macro (Andrew.T.Dyer@macrointernational.com); Geoffrey Halsted, ICF Macro (Geoffrey.W.Halsted@macrointernational.com)

Using computer-assisted methods to collect data has become common in doing survey research in developed countries. The use of computers in personal surveys and telephone surveys (i.e. CAPI, CATI) has facilitated data collection, improved data quality, and allowed for implementation of more complex and innovative questionnaire designs. The performance of real-time methods in developing country contexts is not widely known, however. In this paper, we highlight strengths and challenges we encountered in implementing a CAPI survey in five developing countries.

In fall 2009, the Institute for Health Metrics and Evaluation launched a face-to-face survey in Bangladesh, Indonesia, Peru, South Africa, and Tanzania soliciting opinions from the general public about different health states. The survey is conducted by collaborators in each study site. To accommodate a complex questionnaire design and ensure standardized data collection across diverse sites, we opted to do a CAPI survey with SPSS Dimensions on netbook computers. We designed a system with the ability to synchronize with each netbook in the field, allowing the sites to download updated questionnaires and upload collected data.

An obvious strength with CAPI was the capability of generating a series of randomized questions according to precise weight matrices. Also, it has reduced manual data entry time/costs and has overcome challenges associated with standardizing data entry practices we experienced with previous cross-national PAPI surveys. However, some limitations should be highlighted. Since the internet connectivity in some of these sites is from ideal, downloading updated questionnaires and uploading collected data on a regular basis posed challenges to the sites. We also encountered incidents where interviewers were assaulted and had netbooks stolen, calling for interviewer safety and respondent confidentiality to be reassured. We present summary statistics on data uploading time/frequency, data quality, and interviewer performance and provide recommendations for future CAPI work in developing countries.

Privacy and Confidentiality Concerns of Respondents Who May Have Been Counted Twice in the Census

Jennifer Hunter Childs, U.S. Census Bureau (jennifer.hunter.childs@census.gov); Anissa Sorokin, U.S. Census Bureau (anissa.sorokin@census.gov)

In Census 2000, the U.S. Census Bureau estimated that 5.8 million people were counted twice (Mule, 2002). This could happen for many different reasons, including: a family could move and fill out census forms for both addresses; a child in shared custody could be counted both with his father and his mother; or a couple could be counted at their primary address and at their seasonal home. Ideally, in a census each person should be counted only once. This paper describes two cognitive interviewing studies that explore different ways of resolving duplicates in the census. The goal of both studies was to have the respondent confirm that the person had two addresses, and divulge where he stayed most of the time. This paper compares the privacy and confidentiality concerns that respondents expressed across both studies.

In the first study, cognitive testing was conducted on a small sample of respondents for whom a household member had been identified as a possible duplicate in the Census Dress Rehearsal data (i.e., the same person seems to have been counted in two households). In this study, we told the respondent directly that the person “may have been counted at another residence.” Cognitive testing revealed that, using this approach, respondents may feel that their confidentiality has been breached. In the second study, cognitive testing was conducted on a small sample of respondents for whom it was known by the researchers that a household member had two addresses. This study used a less direct approach that asked a series of in-depth, age-targeted questions about other places the person could have been duplicated. Cognitive testing showed that this approach was less likely to generate confidentiality

concerns, but privacy concerns may still arise, particularly about children. This paper discusses both types of concerns and their implications.

Using Satellite Images for Urban Sampling: The Case of Luanda, Angola

Mark W. Andrews, Survey Consultant (mandrews@cox.net)

In the past few years, high quality and affordable satellite images have become available for most major urban areas in the world. Images that once cost hundreds or even thousands of dollars now can be, in many cases, downloaded for free using popular internet applications. As a result, many who conduct surveys in areas with limited availability of maps and population data have sought to develop methodologies for using these images for sampling.

Luanda, Angola, appeared as an appropriate test case for developing such a methodology. The last census in Luanda was in 1983. Since then it is estimated that population has roughly tripled. The exodus from the countryside during decades of civil war resulted in a city with vast informal slums and unclear boundaries. Detailed maps were not available for all areas.

This paper will detail a satellite-based sampling methodology used in a household study in Luanda. This methodology was designed as an affordable alternative for a project under budget constraints. The paper will discuss the strengths and weaknesses of this methodology relative to other available methods.

Finally, it will discuss possible ways to improve the method used and discuss in what context this methodology could be employed in the future.

Comparative Research

Beliefs and Misbeliefs About the Causes and Treatment of HIV/AIDS in Sub-Saharan Africa

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

(amccutcheon1@unl.edu); Robert Tortora, The Gallup Organization Ltd

(Bob_Tortora@Gallup.com)

The spread of HIV and AIDS throughout Sub-Saharan Africa has created notable alarm among world health officials. Although there is evidence of changing public attitudes and behaviors related to HIV/AIDS in recent years, Sub-Saharan Africa continues to be the worst affected area with an estimated 63% of the world's HIV infected population, and 72% of the world's AIDS deaths (World Bank estimates for 2006). In 2009, the Gallup World Poll asked nationally representative samples of respondents in 30 Sub-Saharan African nations about their beliefs as to behaviors related to HIV/AIDS. In this paper, we use a multi-level latent class model (Vermunt 2007) to examine the nature of the publics' beliefs about these behaviors in these 30 nations. Preliminary analysis of the World Poll data indicates that the populations of most of the 30 Sub-Saharan nations can be characterized as belonging to one of 2 (or 3) classes — those who appear to be relatively well informed, and those who are not. Interestingly, few respondents appear to be in an intermediate (e.g., partially informed) state. The implications of this, as well as an analysis by demographic groups (e.g., gender, age, urban/rural) and national-level variables (e.g., rates of infection, information expenditures) will also be explored using these multi-level models.

Mapping EU Attitudes: Conceptual Considerations and Empirical Evidence

Hajo G. Boomgarden, Amsterdam School of Communications Research

(H.Boomgarden@uva.nl); Andreas R.T. Schuck, Amsterdam School of Communications

Research; Matthijs Elenbaas, Amsterdam School of Communications Research; Claes H. de

Vreese, Amsterdam School of Communications Research

Public attitudes towards European integration and the EU are at the heart of a growing body of research. The conceptualization of these attitudes, however, lacks conceptual and empirical clarity, with some studies referring to euroskepticism and others to EU support, relying on sometimes the same, sometimes different measurement instruments. This study argues that with a growing diversification of the EU's policies, a one-dimensional approach to EU attitudes may be outdated. Furthermore, existing literature by-and-large was oriented towards rather utilitarian oriented attitudes, whereas emotional responses are not addressed. We review existing approaches of conceptualizing and measuring EU public opinion and based on this inventory empirically distinguish different dimensions of EU attitudes. The findings indicate five dimensions of EU attitudes to be differentiated, relating to the notion of a European identity, to the future of European integration in terms of deepening and widening, to affective responses to the EU, to the performance of the EU and its institutions, and to utilitarian considerations. Based in this differentiation, the study furthermore argues that it is inappropriate to expect the same predictors of EU attitudes to equally apply to these different dimensions. We show that some predictors relate more strongly to certain than to other dimensions. The analyses are based on a representative web-survey in the Netherlands in November 2008 (n = 1,390) that includes 25 items measuring EU attitudes and measures of various antecedents of EU attitudes. We discuss the appropriateness of the conceptualization of EU attitudes in prior studies and put forwards that more attention needs to be given to the measurement of EU attitudes and to an appropriate modeling of predictor variables for different dimensions of EU attitudes.

Measuring Economic Performance and Social Progress

**Miguel E Basanez, Cultural Change Institute, Fletcher School, Tufts University
(mbasanez@gmail.com)**

In February 2008, the French President, Nicolas Sarkozy, out of dissatisfaction with the current way of measuring economic performance and social progress, appointed a commission headed by Joseph Stiglitz, Amartya Sen and 25 other prominent social scientists. On September 14, 2009, the commission presented a 292-page report calling for the abandonment of "GDP fetishism". The commission focused on three aspects of the problem: first, criticisms of GDP as a measure of well-being and the justification to replace it; second, a search for measures of "quality of life"; and third, ways of incorporating sustainability, i.e., the well-being of future generations.

In this paper I will argue that with the current knowledge mined by the World Values Survey group and framed by Ronald F. Inglehart, it is possible to produce such measurement of economic performance and social progress based on subjective indicators. However, the challenge that remains is to replicate such an index using objective measures. I hope to present a tentative index of economic performance and social progress based on objective indicators, built along the lines of the UNDP Human Development Index and upon Inglehart's contributions.

What Should Congress Members Do? Using Survey Embedded Experiments to Study Citizens' Clientelistic Expectations in Mexico

Pablo Paras, Data - OPM (pp@dataopm.net)

How do citizens view the work of their legislators? Do they think representatives should legislate and oversee the executive, or do they believe representatives should deliver resources to their community and help individuals with problems? Do they expect both? In Latin America, as in the U.S., traditional surveys typically show that popular evaluations of the national congress are overwhelmingly negative but are not able to explain why. Are congresses held in low esteem because people think that congress is not legislating well or efficiently overseeing the executive, or could it be because citizens do not know what

legislators are supposed to do and want them to deliver pork for their district and personalistic benefits for themselves?

Since clientelism targets the poor and plays an important role in elections in developing democracies, it is important to know if poor citizens truly have different views than wealthier citizens about what a member of congress should do once elected. We use quasi-experimental procedures in a July 2008 public opinion survey to begin to address these questions. Our main hypothesis is that people of lower SES will have more clientelistic expectations of congress members than people of higher SES. Poor people will think that a representative should work to help individuals with problems, obtain resources for their community and maintain contact with the community, rather than legislate, oversee the executive, and work on congressional standing committees. We test this hypothesis using an original dataset of 800 respondents in Mexico City. Our survey incorporates several experimental treatments to explore how people of different SES respond to different types of behavior by congressional candidates and incumbents. Survey questions also allow us to explore whether opinions about the job of congress members differ systematically with SES, and whether they received social welfare benefits from the government.

Election Polling

Once in a Lifetime Voters? "Defining Moments" and Their Impact on Electoral Turn-Out
COLIN MCILHENY, PRICEWATERHOUSECOOPERS (colin.mcilheney@uk.pwc.com)

It is rare to have real-life exact like for like matched pair comparisons where the voter turn-out in an electoral geography with identical boundaries for two contests can be analyzed over such a short period as a matter of weeks.

The paper will focus on two case studies to explore the thesis that what are perceived to be "defining moments" in history have a major impact on voter turn-out.

The chosen case studies are separated by a decade and they cross international boundaries- but they are united by some common characteristics. They are firstly the Northern Ireland 1998 Good Friday Referendum which produced one of the highest turn-outs ever in the UK compared to the election for the Northern Ireland Assembly (at the time thought to be of equal importance) held just a few weeks later. The second example is the 2008 US Presidential/Senate election in Georgia versus the subsequent Senate Run-off race. Access to micro level turn- out statistics helps to aid our understanding in both cases

Why in both the case studies was there such a marked disparity in turn- out between what is widely seen to be the "defining moment" and the subsequent electoral contest? The paper uses the case studies as a vehicle to explore competing theories on motivations for voter turn-out and examines a variety of potential reasons to explain this phenomenon

Should We Use Recall of Previous Vote(s) to Weight Electoral Polls?
Claire Durand, Universite de Montreal (Claire.Durand@umontreal.ca); Isabelle Valois, Universite de Montreal (Isabelle.Valois@umontreal.ca); Mélanie Deslauriers, Universite de Montreal (Melanie.Deslauriers.2@umontreal.ca)

In many European Countries - e.g., France, Norway, Great Britain -- recall of previous vote(s) is used in order to correct for possible sampling bias and reduce sampling variance when estimating electoral

support for political parties. This practice appeared in the sixties to compensate for the then systematic underestimation of the support for the extreme left-wing parties. It has been maintained in many countries until now, while the systematic underestimation of the support for the extreme-right appeared to have replaced the underestimation of the extreme-left. However, according to Waldahl and Aardal (1982), adjustment for declared previous vote tends to further under-represent support for under-represented parties when their electorate is growing and, on the opposite, over-represent it when their electorate is decreasing. This is exactly what happened in the French presidential elections of 2002 (Durand et al., 2004) and 2007 (Durand, 2008). The present study aimed at updating research on this topic. It addressed two main questions: 1) Is recall of previous vote reliable and 2) what is the impact of using this information to adjust data? In order to answer the first question, we used longitudinal data from the Canadian election study panel of 2004-2008 and from polls conducted in Quebec by Crop in March 2007 and December 2008. In order to answer the second question, we used a series of polls conducted in Quebec between the elections of March 2007 and December 2008, as well as the French Electoral Panels conducted for the 2002 and 2007 presidential elections. These data allowed us to examine the relationship between the evolution in declared previous vote and the evolution of support for the different parties. We conclude on whether, when and why adjustments for declared previous vote may lead to bias in the estimation of support for the different political parties.

The 2004 U.S. Exit Poll Controversy Revisited: Why Were Bush Voters Reluctant to Participate in the Exit Polls?

Tamas Bodor, University of Wisconsin-Stevens Point (tbodor@uwsp.edu)

Leaked exit poll data on the 2004 Presidential election suggested that John Kerry was ahead in several key states and was winning the election. Yet the official vote tallies showed that George W. Bush easily secured his second term.

The unprecedented degree of discrepancy between the exit poll estimates and the official tallies fuelled a fierce controversy. Since no particular methodological shortcomings had been identified, some advanced the “reluctant Bush responder” hypothesis, arguing that the exit polls overstated the Kerry vote because Bush supporters were less willing to express their views by participating in the exit polls. Since no information existed on those alleged reluctant Bush supporters who refused the interview, this hypothesis relied on rather fragmented, circumstantial evidence. Also, the scope of the reluctant Bush responder hypothesis was seriously limited: it remained virtually unaccounted why Bush voters would have been reluctant to participate. Due to its circumstantial empirical standing and limited scope, the hypothesis was criticized and deemed infeasible. This criticism had far-reaching implications: in lieu of hands on evidence on what made the exit polls overestimate the Kerry vote, some alleged that not the exit polls but the vote counts might have been inaccurate.

This study scrutinizes the reluctant Bush responder hypothesis in light of the 2004 National Annenberg Election Survey (NAES). The findings constitute evidence that, first, relative to Kerry voters, Bush supporters were less outspoken regarding politics during the last month of the campaign. Moreover, the differing state level public outspokenness rate of the Kerry and Bush voters (derived from NAES data) correlates with the state level exit poll error (within precinct error). Second, the findings shed light on the causal structure of the reluctant Bush responder phenomenon: the public opinion dynamics of the campaign are found to be consistent with the spiral of silence theory.

The Role of Exit-Polls During Elections in Ukraine, 1998-2010

Vladimir I Paniotto, Kiev International Institute of Sociology (paniotto@kmis.kiev.ua); Natalia N Kharchenko, Kiev International Institute of Sociology (nkh@kiis.com.ua)

In the report the role of exit-polls will be considered during elections in Ukraine since 1998, when the first exit-poll has been conducted, up to 2007 (parliamentary elections 1998, 2002, 2006 and 2007, presidential elections 1999 and 2004). The deep public resonance was aroused by exit polls in 2004, they were among the reasons for massive demonstrations (the so-called “Orange Revolution”) and revoking the election results. It is supposed to include also the presidential elections, which will take place in January 2010, in our analysis.

According to WAPOR “ Guidelines for exit polls and election forecasts ”: “ Exit polls can serve three different functions that are not mutually exclusive: predicting election results; describing patterns of voter support for parties, candidates, and issues; and supporting extensive academic research efforts ”.

However, in Ukraine, as well as in some other post-Soviet countries, the main objective of the clients, ordering exit-polls, is validation of election returns.

There were almost no exit-polls which have been ordered by media, the majority of exit-polls have been ordered by political forces. What should do pollsters in such conditions? Whether there are enough recommendations which are given in WAPOR “Guidelines for exit polls ... » or additional recommendations are required?

The aspiration of political forces to affect results of exit-polls leads to occurrence of specific methodical and organizational problems in conducting the researches. In the report we will discuss experience of the decision of such problems in exit-polls, which have been conducted by the Kiev international institute of sociology and by other companies. Factors, which influence accuracy of exit-polls and dynamic of accuracy of exit-polls in Ukraine from 1998 to 2010, will be considered.

The Unpredictability of Pundits and Predictable Elections: Using Public Opinion to Predict Political Disputes

Clifford Alexander Young, Ipsos Public Affairs (clifford.young@ipsos.com)

Every political season the pundits and election wonks come out of the woodwork. They spin their yarn on the most probable outcome. Independently of any one perspective, the collective wisdom typically points towards a political race which will be competitive. The pundit’s basic perspective is that elections are mysterious, unpredictable things.

In this same vein, political spin doctors believe that only if the right thing were said, in the right way, to the right people, at the right time, then their candidate truly would have a shot—much like die-hard sports fans, routing for their team to the bitter end. For them, public opinion is a malleable, movable thing.

In contrast, a scattering of empirical and scientific evidence suggests the exact opposite—that public opinion, and hence outcomes depending on public opinion such as elections, are quite stable and predictable under specific conditions. This evidence comes from three, very different, sources.

So which perspective is correct? Are elections predictable? Or are they mysterious, unpredictable things?

We will attempt to answer the above question in this paper. To do this, we analyze 102 executive branch elections (both presidential and gubernatorial) in 17 countries. Here we do not want to fall into the generalizability trap, hence the large numbers of elections across countries.

It is worth noting that when we talk about “predicting” elections, we are not referring to electoral prediction a day (or a month) prior to the day of election. In such cases, pre-electoral polling already does this to good effect.

Instead, we intend to raise the bar. Our model incorporates public opinion measures (presidential approval ratings) from six months to one year out from election-day. If such longer-term prediction is possible, the applications, then, are manifold.

New Media and Media Use

An Analysis of the Impact of Traditional News and New Media on the Knowledge of Current Events **Thomas B Christie, University of Texas-Arlington (christie@uta.edu)**

As more news consumers throughout the world turn away from traditional sources of news, profound changes have occurred in the way people receive their knowledge of current events. An understanding of these changes is needed because of the linkage of this knowledge to the formation of public opinion and to the democratic process.

Using data drawn from the Pew Research Center April 2008 Media Survey, this study is a secondary analysis revealing factors contributing to public knowledge of national and world events, emphasizing the role of both traditional and new media usage. In this study, new media are defined as "those media technologies, mostly electronic and digital, that are undergoing expansion in our times" (Williams, Rice & Rogers, 1988, p. 3).

The Pew Research Center survey identified four audience segments defining trends in news consumption: Integrators (people who obtain news from both traditional sources and the internet), Net-Newsters (people who primarily use the web for news), Traditionalists (receive news from traditional sources such as the newspaper) and the Disengaged (those with little interest in news) (Pew, 2008). This paper goes beyond the findings of the Pew study to further examine the data in this study, including variables appropriate for ascertaining how traditional and new media usage predicts knowledge of current events. 'Knowledge of current events' as the dependent variable in this study, uses a bank of four questions included in the Pew study. Independent variables include key demographic variables and indicators of newspaper, cable television, radio news, television news, network news, talk radio, newspaper, social media and internet news consumption. A regression analysis reveals the key demographic, traditional news and new media variables that predict knowledge of national and international news and events necessary for public opinion formation.

Digital Literacy and the Gender Divide in Online Information Consumption: User-Sophistication and Susceptibility to Information-Overload **Muzammil Hussain, University of Washington (muzammil@uw.edu)**

More Americans now turn to the Internet for political information than to newspapers, going online to legacy news sources as well as online portals, news aggregators, and political blogs to learn about politics and form political judgments. The rapid increase in the abundance of civic information and audience selectivity has led to normative concerns of attitude-reinforcing selective exposure, selective avoidance of ideologically challenging information, and the potential for information-overload (i.e. cognitive fatigue arising from trying to sort through the abundance of information). However, in addition to gender, age, and SES, what role might digital literacy and technical user-sophistication play in mediating outcomes of information-overload? Attempts to investigate the role of technical literacy in citizens' online experiences have traditionally been hampered with limitations of measuring digital literacy on self-reported measures of familiarity with online activities and various Web objects, such as blogs, browsers,

JPGs, etc. In this study, with data from a national survey, we test for the role of digital literacy, measured as engagement with a robust index of 20 contemporary online activities, in addition to age, gender, and SES, on outcomes of information overload. The results show that gender is a significant predictor of susceptibility to information overload, with females significantly more likely than males to report cognitive fatigue to information abundance. Furthermore, a generational analysis finds that these disparities are most significant for Gen X (18-30), and Boomers (43-61), but not for Gen Y (18-30) or Seniors (62+). This study lends support to the digital natives hypothesis that younger citizens may be better equipped to naturally self-adapt to disparities in digital literacy over time, while this may not be true for older citizens. The implications for a second-level digital divide based on user-sophistication and its potential for differential gains from online news consumption, are also discussed.

How do Country Contexts Influence Media Use?

Joachim Bruess, InterMedia (bruessj@intermedia.org)

This paper examines how and to what extent country contexts such as the level of political freedom and civil liberties or economic performance influence media use at the individual level. A hierarchical linear model will test the hypothesis that country-specific restrictions and limitations are associated with limited individual media use. Data for the analysis are based on nationally representative surveys from more than 60 countries conducted in 2008 and 2009. Variation of individual media use is related to getting news and information from the radio, TV, newspapers, mobile phones and the internet. Insights from this study will provide helpful guidance to media development efforts in a variety of non-Western countries across the world.

Should I Support the Internet Censorship in China? Toward an Integrative Framework of the Theory of Reasoned Action, the Third-Person Perception, the Authoritarianism Personality, and the Unique Characteristics in China

Guangchao Feng, Hong Kong Baptist University (fffchao@gmail.com)

Internet censorship in China has been widely studied since the advent of the Internet in China in the mid-1990s, yet few studies have been done to explore what Chinese people think of the censorship and why they think so. Despite the inconvenience caused by the censorship in general, many Chinese people, especially college students still support the government to block foreign websites. Relevant theories were reviewed and empirically examined to investigate the possible causes. After the right-wing authoritarianism personality measurement scale of Chinese version was developed in study 1, it was used in study 2, in which hundreds of college students in China were queried. Bayesian Path Analysis revealed that the subjective norms particularly the norms from the teachers and parents, the authoritarianism personality especially the submissive personality, and the unique Chinese characteristic, i.e. the single-party state, successfully predicted the behavior of supporting Internet censorship either directly or indirectly, whereas the third-person effect, which is well supported by previous research in explaining why people support censorship after reading objectionable messages, and some other unique Chinese characteristics, such as the Confucius family tradition and the one child policy, failed to play a role. The reasons for those mixed results were discussed, and directions for future research were also suggested.

AAPOR Concurrent Session A

Thursday, May 13

1:30 p.m. - 3:00 p.m.

Address Based Sampling Issues & Challenges.

Investigating Lister Error Associated With Units Added in Dependent Area Listing

Aliza Kwiat, U.S. Census Bureau (aliza.s.kwiat@census.gov)

The U.S. Census Bureau currently uses area listing methods in several programs. Area listing is used to develop the address list for decennial censuses and to develop the rural portion of the demographic household survey frames. Area listing is usually considered to have the best quality among available methods of developing survey frames. However, there has been very little research done to investigate the errors of area listing.

This paper will look into the quality of the units added in a dependent listing by two different listers canvassing the same blocks between May and August 2007. The listers did not know that these listing assignments were different from regular listing assignments. Previous research focused on the units in this listing that were sent out and the findings showed that area listing is not perfect and has room for improvement. A paper detailing those results has been published in the 2009 JSM Proceedings (Kwiat, Aliza 2009. Examining Blocks with Lister Error in Area Listing. JSM Proceedings, Survey Research Methods Section. Alexandria, VA: American Statistical Association.). This research focuses on the units added by the listers during their listing – we do not investigate lister characteristics. Because there were only nine out of the 176 blocks that had complete agreement between the two listers on the units that were added, the results of this listing provide data that is rich in information and will help us understand what challenges are faced by listers in the field. How consistent are their results? What can be said about those units that were only added by one of the listers? Were the units added truly adds to the existing frame? What are some common problems found with the added units? Overall, what is the value to the frame of the units added during area listing?

Your Home Was Specially Selected: Using Address-Based Sampling as a Recruitment Technique

Justin T. Bailey, The Nielsen Company (justin.bailey@nielsen.com); Gretchen Grabowski, The Nielsen Company (gretchen.grabowski@nielsen.com); Michael W Link, The Nielsen Company (michael.link@nielsen.com)

Recent research has shown the effectiveness of address based sampling (ABS) as a sampling frame that provides wider coverage than random digit dialing (RDD), principally because it includes cell phone-only households that are not part of an RDD frame. However, the ability to recruit cell-phone only homes via phone hinges on households supplying a phone number on returned mail materials. Research has shown that compliance with this question on mailed materials has typically been around 50-55% among responders (Link et al., 2009), so there is still a strong need to improve response rates to ABS pre-recruitment materials.

Nielsen recently attempted to expand on previous ABS techniques by using ABS pre-recruitment materials as the first step in the recruitment process. A one-page informational flyer was included with the

pre-recruitment questionnaire that explained how to participate and what the next steps were. Materials were worded in such a way as to generate initial interest in participating. Further, we placed more importance on the acquisition of a phone number than in previous studies, indicating that it would be required to participate.

In another departure from previous ABS methodology, we mailed pre-recruitment materials to both matched and unmatched households to assess pre-recruitment return and subsequent accept rates among the matched group. It was hypothesized that these enhancements would improve response rates at the pre-recruitment phase among both the Matched and Unmatched households, and lead to collection of more phone numbers for unmatched households.

Analyses revealed that these changes to the pre-recruitment materials, compared to the Nielsen diary service, significantly improved return rates and phone number inclusion among unmatched households. This, in turn, improved rates among hard-to-reach demographic groups. These findings and possible explanations are discussed.

Non-Response Bias: Recent Findings From Address-Based Panel

Joe Garrett, Knowledge Networks (jgarrett@knowledge networks.com); J. Michael Dennis, Knowledge Networks (mdennis@knowledge networks.com); Charles DiSogra, Knowledge Networks (cdisogra@knowledge networks.com)

Achieving reasonable levels of response to surveys continues to be problematic as response rates have generally declined over the past several years across numerous survey panels and survey modes. While this certainly is not a favorable trend, it is important to remember that the bias in survey estimates due to such nonresponse is a multiplicative function of two components – the nonresponse rate as well as the overall difference between respondents and nonrespondents. Recently, Knowledge Networks (KN) started to replace its RDD-based panel recruitment for KnowledgePanel® with address-based sampling (ABS) using the U.S. Postal Service Delivery Sequential File. This was done to provide better statistical coverage of cell-phone-only households that now represent about 20% of all U.S. households as well as improve coverage of young adults and minorities. One research effort arising from KN's ABS recruitment strategy was that KN was able to keep a record of those persons who joined the KN panel and those persons who refused to join, and then link the address of the nonresponders to other databases with modeled data for household characteristics as well as county-level data such as 2008 presidential candidate vote counts. This paper presents the results of the various comparisons that were made between recruited and sampled cases. Results of these recruited/non-recruited comparisons across several demographic and socio-economic variables and Obama/McCain vote tallies show householders having a similar propensity to join KnowledgePanel. The findings help reviewers of KnowledgePanel survey data to better understand the potential, however limited, effect on bias due to nonresponse stemming from differences between respondents and nonrespondents at the KN panel recruitment stage.

Address-Based Sampling Design and Coverage Rate: The Case of REACH U.S.

Kanru Xia, NORC (xia-kanru@norc.org); Kari Carris, NORC (Carris-Kari@norc.org); Whitney Murphy, NORC (Murphy-Whitney@norc.org); Hee-Choon Shin, NORC (shinh@uchicago.edu)

For the last several decades random digit dialing (RDD) has been utilized for large-scale surveys. RDD has been popular because the sampling frame contained all U.S. households with landline telephone services. The RDD coverage rate is steadily declining due to recent telephony changes and increases in cell phone-only or cell phone-mostly households. Address-based sampling (ABS), coupled with multiple

modes of data collection such as telephone, mail, and face-to-face components, has emerged as an alternative method to address issues with RDD coverage.

We will examine the resulting coverage rate with the implementation of ABS design for the Racial and Ethnic Approaches to Community Health across the U.S. (REACH U.S.) Risk Factor Survey. REACH U.S. is the cornerstone of CDC's efforts to eliminate racial and ethnic health disparities in health priority areas including breast and cervical cancer, cardiovascular disease, diabetes mellitus, adult/older adult immunization, and hepatitis B. In 2007, CDC established cooperative agreements with 40 organizations for each to develop and implement innovative strategies to reduce health disparities among African Americans, Native Americans, Hispanic Americans, Asian Americans, and Pacific Islanders. NORC is conducting surveys in 28 REACH U.S. communities across the country.

REACH U.S. employed an ABS approach combined with data collection via telephone, mail, and face-to-face interviews to obtain the requisite number of completes in each community. The mailed survey included a question to assess whether the household had access to cellular phones, landline phones, or no telephone service. We will examine the pattern of telephone service among responses to the mailed survey to determine the extent to which cell phone-only or cell phone-mostly households are covered by the mail survey. Preliminary results indicate that a substantial proportion of cell-only or cell-mostly households are covered by mail.

Undercoverage Error in List-Assisted RDD Surveys: Evaluation and Adjustment
Burton Levine, RTI International (blevine@rti.org)

Two recent phenomena have threatened coverage properties of established methods for telephone surveys: individuals with only cell phones and degraded properties of the landline frames. Coverage bias from the former has been demonstrated and addressed through the use of dual landline and cell phone frame designs, while coverage bias from the latter has not been evaluated and remedies have not been developed despite the potential for error.

List-assisted random-digit-dialing (RDD) improves sample efficiency as fewer nonworking numbers are dialed but introduces undercoverage of telephone numbers that are (usually) in groups of 100 telephone numbers with no listed numbers (zero-listed 100-banks). Recent empirical research places the proportion of landline numbers in zero-listed banks to be somewhere between 5% and 20%, yet the effect on coverage bias remains unevaluated and there are no available methods for its adjustment.

Evidence shows decreasing listed numbers in 100-banks and a growing number of zero banks that contain unlisted working numbers. Households in banks with few listed numbers are at threat of dropping from the list-assisted RDD frame and are likely similar to those already in zero-listed banks. Each telephone number has a probability of being in a zero-listed bank, and the count of listed numbers in a bank is a key determinant. We use this premise to achieve two goals: estimate coverage bias due to omission of zero-listed banks and describe coverage adjustments to compensate for it.

Coverage bias is estimated using linear associations between bank counts and survey estimates, groups based on bank counts, and logistic propensity models employing additional variables. Estimates are produced based on data from three RDD studies, each with over 10,000 respondents. Estimates of undercoverage due to zero-listed banks are also compared to direct estimates of undercoverage due to exclusion of cell phones. Methodology for weighting is proposed and discussed.

Assessing Public Opinion in War Zones

Afghanistan: Where Things Stand

Gary Langer, ABC News (gary.langer@abc.com); Matthew Warshaw, D3 Systems and Afghan Center for Socio-Economic and Opinion Research (matthew.warshaw@d3systems.com)

More than a hundred trained Afghan interviewers and field supervisors set to work this December on the fifth national public opinion poll in Afghanistan sponsored by ABC News and international media partners, with field work directed by the Afghan Center for Socio-economic and Opinion Research (ACSOR) in Kabul. Following random sampling procedures, field teams conducted in-person interviews of 1,534 Afghan adults across all 34 of the country's provinces. Results showed sharp gains in views of the country's direction and support for its president alike, bolstered by progress in countering the Taliban, improved development efforts and economic advances. However significant challenges remain: Complaints about official corruption are higher than ever, views of the United States and NATO's performance remain poor, accounts of local violence have held steady and the corrosive effect of civilian casualties remains a threat to support for the presence of ISAF forces. Our presentation will include a description of field work and discussion of the challenges of survey research in these conditions as well as a summary of key findings, with trends over time.

Challenges of Including Cell Phone Households in Cross-national RDD Surveys

A Review of Developments and Associated Challenges With the Inclusion of Cell Phone Households in Cross National RDD Surveys

Robert Manchin, Gallup Europe (manchin@gallup-europe.be); Bob Tortora, Gallup London (bob_tortora@gallup.com); Rajesh Srinivasan, Gallup Organization (rajesh_srinivasan@gallup.com); Femke De Keulenaer, Gallup EUrope (femke_de_Keulenaer@gallup-europe.be); Gergely Hideg, Gallup Hungary (gergely_hideg@gallup.hu); Marek Fuchs, Institut für Soziologie Technische Universität Darmstadt (fuchs@ifs.tu-darmstadt.de); Vasja Vehovar, University of Ljubljana (vasja.vehovar@fdv.uni-lj.si); Allan McCutcheon, UNL (allan_mccutcheon@gallup.com)

In most countries around the globe the penetration of cell phones has reached levels similar to that of landlines before RDD surveys became prevalent. Evidence from different countries suggests that those who can be reached primarily via cell phones, even if they also have household level access to fixed line phones, differ on many aspects from those who are primarily reachable via household fixed lines. Given the rapidly growing rates of cell phone households both within the developed and more importantly in the developing world, where the penetration of landline households is declining, designing future phone-based survey research requires an understanding of the various available options. Members of the panel have been involved in on-going large scale mobile phone multi-country surveys, like the Flash Eurobarometer and the Gallup WorldPoll. They have accumulated research experience related to the challenges of developing and executing national RDD designs with optimal properties such as high coverage, reasonable costs and high efficiency.

This panel will bring together a wide range of survey practitioners from different parts of the globe who discuss their experiences in developing and executing RDD designs and the challenges they encountered while incorporating cell phones. They address developments in Europe and in parts of the developing

world that present opportunities and challenges for conducting survey research over cell phones in the future.

Sharing of Mobile Phones – Consequences for Sampling and Weighting

Marek Fuchs, Darmstadt University of Technology (fuchs@ifs.tu-darmstadt.de); Britta Busse, Darmstadt University of Technology (busse@ifs.tu-darmstadt.de)

With declining landline telephone penetration rates and increasing mobile-only rates it has become questionable whether it is still feasible to conduct telephone surveys in the fixed-line frame only. In fact it seems advisable to combine the fixed-line frame with the mobile phone frame to counteract the risk of a growing coverage error. The dual frame sampling approach offers a statistical framework for this intent. For calculating a person's selection probability in a dual frame survey the number of fixed-line phone numbers of the household where the person is living is considered as well as the number of eligible persons who could answer the phone. In addition, the number of mobile phone numbers is regarded in the computation of the selection probabilities. This approach is reasonable since mobile phones are considered personal devices that would not be shared among groups of people. However, recent studies in the US and Europe have demonstrated that mobile phone users might in fact answer incoming calls on other people's mobile phones (active sharing) and that other people answer calls on the respondent's mobile phone (passive sharing). Based on data from the first three waves of our ongoing Mobile Phone Panel conducted in Germany (n=1,200) we will identify household types and individuals who are especially prone to sharing their mobile phones. Further on, we will introduce an adjustment of the formula for the selection probability when using dual frame samples. In addition, we will discuss results with respect to potential respondent selection procedures when multiple eligible persons might answer a survey request over single mobile phone.

Differential Call Scheduling in Dual Frame Telephone Surveys: Should We Be Concerned About "Timing Effects"?

Femke De Keulenaer, Gallup Europe (femke_de_keulenaer@gallup-europe.be); Robert Manchin, Gallup Europe (robert_manchin@gallup-europe.be)

It is well established that time-of-day and day-of-the-week are important factors to consider when calls are scheduled in "traditional" landline telephone surveys (e.g. calls in the evening are more productive). The switch from landline surveys to dual frame surveys – combining cell phone and landline interviews – raises two important questions about optimal call scheduling for telephone surveys.

(a) Calling protocols and respondent availability. Calling schedules for landline surveys are generally skewed towards evening and weekend calling since this is when most people are available at home. Research shows that many people have their cell phones on all day; as such, cell phones provide a direct communication link to respondents at virtually any time. Nevertheless, most cell phone surveys use the same "skewed" calling schedules that are typical of landline surveys. The first question that we want to answer in this study is how call scheduling in cell phone surveys can be optimized in order to increase contact and completion rates.

(b) Measurement bias linked to call scheduling. Research shows that the day and the time that a call is attempted influences nonresponse bias; however, an area that received little attention is the effect that differential calling schedules might have on measurement bias in survey estimates. The second question that we want to address is whether interviews completed during the day produce different estimates than those completed at weekends or evenings (after controlling for individual characteristics of the

respondents and the interviewers), or in other words: should we be concerned about "timing effects" when developing calling schedules for cell phone interviews in dual frame surveys?

We will endeavor to answer these questions by analyzing contact history data of several landline, cell phone and dual frame surveys conducted by Gallup in the past five years in many European countries.

Study of the Dynamics of Cell-Phone Ownership and Usage Patterns and Its Application to the Design of Dual-Frame Methods

Rajesh Srinivasan, Gallup (rajesh_srinivasan@gallup.com); Manas Chattopadhyay, Gallup (manas_chattopadhyay@gallup.com); Steve Hanway, Gallup (steve_hanway@gallup.com)

Over the last few years, there has been a significant increase in awareness and use of dual-frame RDD methods to address the growing cell-phone only population and its impact on coverage in traditional RDD surveys. However, with changes in cell phone penetration and usage patterns across the country, what was state of the art in design and execution six months ago, may be rapidly changing. The Gallup G1K, a nightly survey of 1000 random households across all 50 states uses a dual-frame methodology. The frequency and size of the sample provides us a unique opportunity to track the dynamics of cell phone usage and profile of users both on demographic as well as attitudinal/behavioral measures. In this paper, we will analyze data over a six month period and evaluate the changes from three different perspectives – coverage, costs and efficiency. Specifically, we will review and contrast multiple methods for combining the dual frames and its effect on resulting costs and efficiency.

Cross-National Research on Public Opinion

Latent Support for Sub-National Military Groups in Pakistan

Tarek Albaghal, University of Nebraska (mtalbaghal@hotmail.com); Lei Huang, University of Nebraska (leihuang03@hotmail.com); Ana Lucia Cordova Cazar, University of Nebraska (analucordova@yahoo.com); Allan McCutcheon, University of Nebraska (AMcCutcheon1@unl.edu); Rene Bautista, University of Nebraska (rene.bautista@huskers.unl.edu)

Current world events have focused attention on the Islamic world, and in particular, Pakistan. Sub-national military groups, and support for such groups, have become a prevalent part of this discussion. Prior research indicates that these groups have some, though limited, support from the population of various nations. This research, however, has been unable to identify the complete extent of support for these groups and thus the possible population these groups can recruit from or rely on for popular support. The present research seeks to address this gap. Using a unique data source—the Gallup World Poll data—public opinion in Pakistan was analyzed to assess the level of support for these sub-national groups. As these groups' ideologies are religious in tenor, opinions that are consistent with these religious positions as well as support for the actions of these groups are examined. The Gallup World Poll contains a number of questions indicating the potential support for these groups and the positions they take. Beyond examining the distributions of these individual questions, support is conceptualized as a latent property, measured by a subset of these responses. Support is related to religious intolerance, religious fundamentalism, and approval of sub-national military actions, including past actions such as those on Sept. 11, 2001. Latent supporters are estimated to be a small, yet substantively important, proportion of these populations. This indicates that although the majority of people in these countries do not support these groups, there remains a sizable resource these groups can rely and draw upon. Preliminary analysis estimates that approximately 4% of the accessible populations are strong supporters; about 45%

of the population is neither strong supporters nor opponents, while approximately half of the population is strong opponents of these groups.

Inequality of National Exam and Participation in Higher Education in China- Results from a Pilot Online Deliberative Poll

Rui Wang, Stanford University (wrui@stanford.edu); Alice Siu, Stanford University (asiu@stanford.edu); Qianfan Zhang, Peking University (qfz@pku.edu.cn); Xiaobo Mou, Peking University (mouxbo@126.com); Mai Cheng, Peking University (chengmai@pku.edu.cn)

There has been a long-time debate about the admission system of the national universities and the policy of the national exam in China. Since the universities admit students not based on a uniform score in the national university entrance exam, but based on a predetermined quota for each major in each province, the competition is extremely rigorous in some provinces because only the top one or two students could have the opportunity to be admitted to the top universities in China. This study is the first national on-line Deliberative Poll ever conducted in China to explore people's knowledge of the national exam policies and their opinions about the inequality of the national exam and the university admission system. A national convenience sample, representing 31 provinces, was recruited by posting ads on a popular Chinese website. Participants filled out two questionnaires before and after participating in small-group deliberations by using text-based QQ messengers and watching on-line expert panel discussions. This paper shows how people gained their knowledge about the national exam and the university admission system after obtaining information from the briefing materials, the discussions with peer participants, and the expert panels. In addition, it also shows how people's opinions changed after becoming more informed.

The Life Opportunities Survey, Office for National Statistics UK
Tom Howe, Office for National Statistics (thomas.howe@ons.gov.uk)

The Life Opportunities Survey (LOS) was launched across Great Britain in June 2009. The survey aims to collect information about people's experiences of taking part in a wide range of activities, including employment, education, leisure and use of public services. The survey will provide both cross sectional and longitudinal data on people's opportunities and why they are limited in any of the activities they'd like to take part in.

The LOS has a significant focus on the life opportunities of disabled people, but can be used to analyze opportunities by a variety of equality strands. The survey was commissioned by the Office for Disability Issues (ODI) to provide a benchmark of progress towards achieving ODI's Equality 2025 objective; whereby disabled and non-disabled people should have equal opportunities. Therefore it has been vital to ensure that the LOS was designed to make the survey accessible to disabled people. ONS have designed fieldwork materials to be accessible to disabled people, for example using large text and Braille showcards.

In addition, the LOS will sub-sample respondents who are unable to complete the national survey, due to lacking mental capacity or having cognitive difficulties, to complete a conversational interview tailored to their individual requirements. This conversational interview will address their access to goods and services as well as the choice and control they have over their lives. Some of the respondents to the conversational interviews will be included in observation studies to further our understanding of their experiences.

ONS have worked closely with a Reference Network of 60 disabled people and organizations to ensure that the survey is accessible for disabled people and covers topics that are relevant to them.

Attitude Factors in the Search for Israeli-Palestinian Peace

**Alvin Richman, Retired (Former Senior Opinion Analyst in U.S. State Department)
(RichmanAl@gmail.com)**

This paper examines recent surveys of the Israeli and Palestinian publics regarding (1) underlying conditions affecting their negotiations and (2) identification of mutually acceptable terms for resolving specific issues in dispute. Among the basic attitudinal conditions examined are the impact of the Fatah-Hamas divide on Israeli willingness to make concessions and Palestinian ability to deliver, high levels of mutual distrust and low expectations among both Israelis and Palestinians about directly negotiating a peace accord, and, therefore, reliance by both sides on an active U.S. role to overcome these impediments to achieving their main objectives -- long-term security for Israel and a viable state for Palestinians.

The second objective relies mainly on two recent sets of dual sample surveys -- by the Truman Research Institute and Peace Polls -- each containing identical wording for its Israeli and Palestinian respondents. These provide powerful tools for comparing their views regarding the priority each side gives to eight different issues in the negotiations and the acceptability of various proposals to resolve each of these issues. This enables us to identify those issues in the Israeli-Palestinian dispute that are likely to be the toughest to resolve (e.g., the future of Jerusalem) and which proposals to resolve the various issues have the backing of both publics.

In addition to recent surveys conducted simultaneously in Israel and the Palestinian Territories using identical question wording with each public (Peace Polls, Truman Institute and the Pew Center's "Global Attitudes Project"), the analyses will draw from a number of 2009-2010 surveys based solely in Israel or the Palestine Territories.

Acceptance of GPS Technology in the Diverse Environment of Jerusalem, Israel

Julie L Paasche, NuStats (jpaasche@nustats.com)

Travel Demand Forecasting Models are used by transportation planning agencies to model current transportation system usage and future system demand. Standard data inputs for these models are obtained from household travel surveys (HTS) in which representative households are recruited to complete travel diaries and report travel information during a computer-assisted telephone interview. Respondent burden associated with this collection method is thought to be substantial, resulting in unit and item non-response as well as inconsistent data (Zmud and Arce, 1997; Zmud, 2001). Since the implementation of the Global Positioning System (GPS) in the mid 1990's, researchers have held hope that the passive collection of travel information via GPS could be used to reduce respondent burden, and improve the quality, quantity, and comprehensiveness of model input data. Inexpensive, wearable GPS logging devices are now available making it is possible to conduct a full-scale HTS using person-based GPS capture. Today, the biggest "unknown" in the application of full-scale GPS-assisted household travel surveys relates to respondent acceptance of this technology. While previous studies have documented the characteristics of respondents who accept or do not accept GPS equipment, there is no unified theory of respondent acceptance to inform or predict the diffusion of this technology as the technology continues to improve and the cost continues to decrease (Bricka, 2008; Wolf, et al., 2006).

This paper explores the factors affecting respondents' attitudes toward technology in general, and GPS technology specifically, within the diverse cultural framework of Jerusalem, Israel. Also explored are the complexities of data collection with the specific distinct populations in Jerusalem of non-Orthodox Jewish, ultra-Orthodox Jewish and Arab, as well as the challenges of fielding in multiple languages (Hebrew, Russian and Arabic).

Election Poll Forecasting & Methodological Considerations

Forecasting Elections: Voting Intentions vs. Expectations

**David Rothschild, University of Pennsylvania's Wharton School of Business
(rothschdm@wharton.upenn.edu)**

Joint work with Justin Wolfers ... Our paper compares the accuracy and informational content of election forecasts based on two different polling questions: the "intent question" (i.e. who will you vote for in this election?) and the "forecast question" (i.e. who do you think will win this election?). Intent question data is ubiquitous, because it implicitly translates directly into one of two main outcome variables, the expected popular vote share of the candidates. Yet, the only information in the response is the intent of the respondents. Forecast question data, while similar to the second main outcome variable of probability of victory, does not directly translate into anything. Thus, while it is occasionally asked in polls, it is rarely utilized. Yet, its response provides information about the intent of the respondents as well as the intent of voters in the respondents' communities and common knowledge about the election within those communities. Optimal transformations of the data, 932 unique polls where both the intent and forecast question were asked, demonstrate that the forecast question provides more valuable information for the outcome variables than the more commonly quoted intent question. A combination of both questions is the strongest forecast. Further, a model is created to determine the information that goes into the response to the forecast question, so we can estimate the percentage of respondents who split their answers (i.e. forecast victory for the candidate for whom they do not intend to vote). We determine that using data on the splits, combined with the intent and forecast responses, provides the most efficient information for accurately forecasting the two outcome variables. We care about providing accurate forecasts, because business and politics both rely on a firm understanding of the underlying values at any given point.

Weight a Little and See if Things Change: The Impact of Weighting Public Opinion Data by Party Registration

Andrew E Smith, UNH Survey Center (andrew.smith@unh.edu); David W Moore, UNH Carsey Institute (dmoore626@comcast.net)

One of the more frequent criticisms of public opinion polling is that Democrats, or Republicans, have been under-sampled, thereby making the results of the particular survey misleading, at best. This criticism typically comes from partisan commentators, and often carries the implication that the pollster involved did this intentionally in order to impact public opinion. There have been modest research into the value of weighting by party identification (Abramowitz, 2006; Kaminska and Barnes, 2009; and several political bloggers, for example, "Mystery Pollster" Mark Blumenthal, maintain a series of posts both for and against weighting by party ID), but there has been no research to date that has focused on weighting by party registration, rather than party identification. There are indeed concerns with weighting by party identification, most notably that party identification is an attitude that is not fixed but changes with demographic shifts and political events. Party registration should be much less malleable as it records a specific behavior rather than an attitude, and there is a known weighting target as these data are kept and

regularly updated by Secretaries of State.

This paper examines the impact of weighting by party registration in 5 statewide elections in New Hampshire to test whether weighting by party identification improves election predictions. In addition, we examine more than 8 years of quarterly data to judge the impact of weighting by party registration on common public opinion measures such as presidential approval ratings, gubernatorial approval ratings, and direction of the country. Data come from the Granite State Poll, an RDD survey of New Hampshire adults conducted by the UNH Survey Center and from pre-election polling conducted by the UNH Survey Center for media partners.

Why Do Polls Go Wrong Sometimes?

Claire Durand, Université de Montréal (Claire.Durand@umontreal.ca); Martial Foucault, Université de Montréal (Martial.Foucault@umontreal.ca); John Goyder, University of Waterloo (jgoyder@watarts.uwaterloo.ca); Mélanie Deslauriers, Université de Montréal (Melanie.Deslauriers.2@umontreal.ca)

Numerous studies have been conducted to examine specific instances where polls have mispredicted election results i.e., cases where close to all the polls conducted during an electoral campaign have been biased in the same direction. However, no study to date seems to have examined globally whether there are common characteristics of these instances. If these common features exist and we can find them, we may be able to preview situations where polls may be biased and even the likely direction of the bias. Characteristics that may be related to errors in the polls may be methodological as well as socio-political. For this research, using aapornet, wapornet as well as sociological and methodological data bases, we identified more than 40 cases in 15 countries where polls have gone wrong in 44 articles from 1951 to 2008. All these cases have been coded on a number of indicators : type and level of error, methodology used for sampling and weighting, type of election, characteristics of the electoral campaign, conclusions as to the causes of errors. The paper will present the results of a meta-analysis of these cases. Our aim is to uncover new regularities that could explain why polls go wrong in some cases and not in others, regularities that can hardly emerge when looking at case studies.

Statistical Properties of Internet and Telephone Polls in the 2008 U.S. Elections

Douglas Rivers, Stanford University (doug@yougov.com)

The bias and variance of survey estimates in the 2008 U.S. elections is evaluated. It is shown that reported variance estimates are inaccurate, but that alternative estimates perform well. The nominal confidence levels tend to be overstated due to bias in the estimates. However, conditional upon turnout and demographics, selection is nearly ignorable, whether the sample is RDD, area probability sampling, or from an opt-in Internet panel.

A Synthesis and Analysis of Economic, Stability, and Opinion Indicators as Predictors of Electoral Outcomes With Particular Attention to State-Level Elections.

Chris Wilson, Wilson Research Strategies (cwilson@w-r-s.com); Bryon Allen, Wilson Research Strategies (ballen@w-r-s.com)

This paper explores the relative importance of economic and public opinion indicators in assessing the political environment with particular reference to the public opinion indicators that are most useful for assessing the political environment as it affects state-level races.

Our analysis extends and synthesizes the existing literature on the effects of economic and national/international stability indicators on electoral outcomes and the literature on public opinion indicators of the political environment by:

- Simultaneously testing economic and stability indicators and public opinion indicators as explanatory variables for political change.
- Extending this analysis to explore the relative importance of national and state-level indicators as predictors of state-level outcomes.

We assembled twenty years of election outcomes across all 50 states including statewide office, federal office, and state legislature (net change in seats) data along with a parallel data set of economic indicators and public opinion data at the state and national levels.

Preliminary analysis of this data set suggests several important findings:

- Opinion about the national economy is a more powerful predictor of electoral outcomes than either the reality of the economy or state specific opinions, even for state races.
- National opinions of the two parties tend to be on par with or more important predictors of state level outcomes than state indicators overall, but there are important regional differences in this effect.
- The exception to all of these findings occurs when there is an important state-level scandal (such as the pay raise scandal in Pennsylvania in 2006). This kind of state-specific super-issue can “drown out” the national political environment.

The important implication of this analysis for practitioners the strong suggestion that even when polling at the state level we should include a battery of key national perception questions to help understand the environment that will determine electoral outcomes.

Lessons from Multilingual Cognitive Interviews

Strategies for Choosing the Number and Nationality of Spanish-Speaking Immigrant Respondents for Cognitive Testing Studies in the United States

Patricia L Goerman, U.S. Census Bureau (patricia.l.goerman@census.gov)

At the beginning of every cognitive testing project, researchers need to address certain methodological questions. Two of the most basic questions are: a) How many respondents should be included?; and b) What characteristics should they have? These questions can be particularly difficult to answer when dealing with a non-English speaking group comprised of individuals from many different national origin backgrounds, such as Spanish-speaking immigrants in the United States. This paper will examine strategies for choosing the number and type of Spanish-speaking respondents in a series of cognitive interview projects conducted at the U.S. Census Bureau. The projects examined in this paper are related to the Decennial Census, which includes only basic demographic questions, and the American Community Survey, which is a much more comprehensive questionnaire that asks about detailed person and housing characteristics. The paper will discuss how numbers and national origin groups of Spanish-speaking respondents were chosen for each project, whether this number seemed sufficient, and why or why not. The paper will also examine the type of subject matter tested in each study and how this can influence the number and different types of respondents needed. On the whole, it seems that less national origin diversity amongst respondents may be necessary when testing basic demographic questions than when testing questions related to specialized subject areas such as heating and plumbing or country-specific systems such as educational level. With more specialized topic areas it is often more likely that different terminology or different customs are at play across Spanish-speaking countries. The

paper concludes with a plan for more systematic research to help determine the optimal number of Spanish-speaking immigrant respondents for different types of cognitive testing studies in the United States.

Respondent Recruitment, Interviewing and Training: Lessons Learned from a Spanish Language Cognitive Interviewing Project

M. Mandy Sha, RTI International (msha@rti.org); Georgina McAvinchey, RTI International (McAvinchey@rti.org); Leticia Reed, RTI International (lreed@rti.org); Sonia Rodriguez, RTI International (srodriguez@rti.org); George R Carter, III, U.S. Census Bureau (george.r.carter.III@census.gov)

Cognitive interviewing in Spanish is a developing method. As the demand for quality Spanish translations of survey instruments increases, it becomes a method of growing utility and importance. However, prior experiences on respondent recruitment, interviewing, and interviewer training are largely based on pretesting English instruments. This paper discusses the lessons learned from a project undertaken by the U.S. Census Bureau to cognitively test the Spanish language translation of selected questions in a survey. We draw upon findings from RTI interviewer debriefings and observations. Results and discussion focus on the lessons learned, recommendations, and implications for recruiting Hispanic participants with limited English proficiency; administering cognitive interview probes in Spanish; and the logistics of cognitive interviews and training interviewers to conduct them. This research is of interest to methodologists who use the cognitive interviewing methodology to inform questionnaire design.

Comprehension Probes in Multilingual Cognitive Interviews

Virginia Yelei Wake, Statistical Research Division, U.S. Census Bureau (virginia.wake@census.gov)

The increasing number of immigrants and diversity of the American population makes it important to collect high-quality, multilingual survey data. Cognitive interviewing is a tool often used to ensure the quality of data collection procedures. However, we know from past research that cognitive testing techniques that are well-established in English do not always elicit the desired information effectively when translated into other languages (Pan, 2004; Pan et al. 2008).

The purpose of this study is to identify the source of the problem from a sociolinguistic approach. It takes a close examination of the effectiveness of the comprehension probe – one of the major types of cognitive interviewing probes – in Chinese cognitive interviews. The data for this study are from 48 Chinese cognitive interviews conducted across two research projects that pretested translated materials from a U.S. Census Bureau survey. The respondents are Chinese immigrants with a variety of backgrounds, including age, gender, education, and length of stay in the U.S.

This study examines multiple levels of meaning and functions of comprehension probes. It analyzes the external and the internal contexts where the probes are situated and the resulting interactional meaning. Further, the conventional Chinese interpretation of the probes is examined, with particular attention to the Chinese responses that are derived from the standard answers given to these probes.

Our findings show that comprehension probes do not have the same function across all languages and cultures. While these probes are designed to seek the respondents' literal interpretation of the materials being tested, the use of the probes is conventionally interpreted by Chinese respondents as seeking for the implied social meaning of the tested materials. This indicates that these probes have to be

reconstructed with socio-cultural elements in order to collect quality data. Finally, a proposal to address this issue is presented.

How Hispanics Rate Their Health Status: Evidence From Cognitive Interviews

**Alisu Schoua-Glusberg, Research Support Services (alisu@researchsupportservices.com);
Marjorie Hinsdale-Shouse, RTI International (mhs@rti.org)**

The widely used question, which asks respondents to rate their general health status as excellent, very good, good, fair or poor, has proved to be a strong predictor of mortality across populations. However, much of the current literature documents that Hispanic respondents tend to answer this question in ways that differ from non-Hispanics. General findings are that Hispanics are more likely to report fair or poor health than non-Hispanics. Studies have looked at different possible explanations for this phenomenon, from disparate socio-economic levels to age or gender differences, to levels of acculturation in a society.

In this paper we analyze cognitive interview responses to the general health status question as well as the answers to probes about how respondents interpreted the question and chose their health status rating. Perspectives considered in the analysis include cultural views of what aspects are seen as part of overall health (physical, emotional, spiritual), how respondents decide to choose some responses over others, and the apparent reluctance to report excellent or very good health when not recently seen by a doctor.

The 68 interviews were conducted in 2004-2005 to validate the English and Spanish versions of the Latino/Hispanic Adult Tobacco Survey. Two thirds of the interviews were conducted in Spanish and the remainder in English. Respondents from a variety of national origins in Latin America, different age groups and general health answered the questions in their dominant language.

Methodological Briefs: Improving Survey Questions and Response Options

Are Cognitive Testing Results Reliable? An Empirical Investigation

Gordon B. Willis, National Cancer Institute (willisg@mail.nih.gov)

Cognitive testing is commonly used to pretest and evaluate a wide variety of survey questionnaires. However, a key issue in survey pretesting concerns the reliability of the results of cognitive interviews. The current study endeavored to evaluate this evaluation method, by conducting an experimental test. Four cognitive labs independently conducted a total of 148 interviews of English, Spanish, Chinese, and Korean-speaking groups, using a self-administered questionnaire on cancer risk factors. The research question posed was whether the results of this testing would point to similar problems across laboratories and participant groups; a set of different problems; or even contradictory problems. Overall, the results clearly indicated the presence of a single, serious defect in the questionnaire, across all labs and groups. The implications of this finding for the conduct of cognitive interviewing will be discussed.

An Experiment in Measuring Health Insurance Status and Type

Dianne Rucinski, Institute for Health Research and Policy (drucin@uic.edu); Kelly Daley, Abt SRBI (k.daley@srbi.com); Julie Pacer, Abt SRBI (j.pacer@srbi.com)

Among policy and health services researchers it is thought that survey estimates of Medicaid enrollment are biased due to underreporting of enrollment (Lewis, Ellwood, & Czajka, 1998; Blumberg and Cynamon, 1999; Kincheloe, Brown, Frates, Call, Yen and Watkins, 2006; Wheaton, 2007; Pascale, Roemer, and

Resnick, 2009; Davern, Klerman, Baugh, Call, and Greenberg, 2008). The reasons suggested for Medicaid survey report-administrative record discrepancies include survey design features, program attributes and survey respondent characteristics. This study reports the results from a field test of a new set of health insurance measures. An RDD list-based sample of parents/caregivers of children receiving public medical assistance will be asked one of two set of questions about their children's health insurance. Respondents will be randomly assigned to: a) a standard version or b) an experimental version that explicitly asks all respondents who report no coverage if they have one of the government sponsored programs (i.e., "Are you covered by All Kids, FamilyCare, Public Aid, Medicaid, Medicare, coverage through the VA or Veterans' Administration, Medical Assistance, or any other kind of government assistance program that helps pay for health care?" And "Do you get a medical card from the state health or public aid or Medicaid office so you can go to the doctor or hospital?") It is hypothesized that the experimental series that specifically ask about publicly sponsored coverage using a broad range of program labels and about a medical card from a state or public aid or Medicaid agency will produce fewer reports discrepant with administrative records than the standard set of questions. Specifically, it is expected that respondents asked the new experimental series will be less likely report their children are uninsured, or covered by a private plan, or by another government plan than respondents who receive the standard series.

A New Method for Evaluating the Propensity to Guess in Political Knowledge Quizzes

Nick Allum, University of Essex (nallum@essex.ac.uk); Patrick Sturgis, University of Southampton (p.sturgis@soton.ac.uk); Patten Smith, Ipsos-MORI (patten.smith@ipsos-mori.com)

Citizen competence is a concept used by political scientists to understand voting behaviour, political attitudes and engagement with politics. A key indicator of citizens' competence in political life is the extent to which they have a good understanding of how political institutions function and are familiar with the issue positions of the major parties. In other words, how knowledgeable they are about politics. Typically, political knowledge has been measured with the use of true/false/don't know survey items. However, there have been notable criticisms of this format (Mondak and Anderson 2004; Sturgis et al 2008) because of the putative propensity for some kinds of survey respondents to hazard a guess in preference to saying that they don't know the answer. If this is true, it may be that the level of political knowledge is underestimated for groups that are dispositionally less likely to guess. In this paper we use data from a British national telephone sample collected by Ipsos-MORI to examine this topic. Our method is to pose political knowledge questions where the allowable response options are all false. This provides us with a relatively pure indicator of the propensity to guess. We evaluate the extent of guessing in relation to key sociodemographic and personality variables. We also estimate 'true' levels of political knowledge by asking other quiz questions with open ended response formats. This enables us to find out whether citizens who are knowledgeable about politics are more likely than others to guess the answers to survey items when they do not know the correct answer.

Illuminating Optimal Survey Response Using Follow-Up Questions

Linda K Owens, Survey Research Laboratory, University of Illinois, Chicago (lindao@srl.uic.edu); Sowmya Anand, Survey Research Lab, University of Illinois (sowmya@srl.uic.edu)

The four stages involved in answering a survey question optimally are: understanding the question, retrieving relevant information from memory, combining the information into a summary judgment, and mapping that judgment onto one of the scale points offered (Tourangeau, 1984). When we analyze ratings obtained on a survey, we have information only about the fourth step of this process. We do not have information on the specific judgments that went into the rating. One way of accessing these judgments is by asking respondents a follow-up question as to why they picked the scale point they did.

The purpose of this study is to examine the consistency between the scale point chosen by respondents and the reasons they specify for doing so. Using data from two surveys of participants' experiences during summer school programs, we compare specific ratings of aspects of the program evaluated on a 5-point scale to the qualitative comments offered as reasons for choosing those scales. On each survey, participants were asked to evaluate the room in which the instruction took place as well as the slides and handouts provided. They were also asked why they evaluated these aspects the way they did. We evaluate the verbatim comments based on: (1) the number of positive or negative characteristics they mention, and (2) the number of positive or negative descriptors they use. Two raters independently coded the open-ended comments using these criteria. With this analysis, we attempt to assess the degree of consistency between steps 3 and 4 of the model, and the utility of follow-up questions in illuminating the cognitive process involved in respondents' choice of a scale point.

Forced-Choice vs. Yes-No Questions: Data Quality and Administrative Effort

Sowmya Anand, Survey Research Laboratory, University of Illinois (sowmya@srl.uic.edu); Linda K. Owens, Survey Research Laboratory, University of Illinois (lindao@srl.uic.edu); Jennifer A. Parsons, Survey Research Laboratory, University of Illinois (jparsons@srl.uic.edu)

The literature recommends the use of forced-choice questions over yes-no type questions, because the former are less susceptible to acquiescence bias and therefore yield better quality data, including better reliability and validity. However, forced choice questions violate norms of everyday conversation and therefore may affect how respondents react to them. Also, because they involve more words, they take longer to administer, adding to the overall length, and therefore cost, of the interview. In an RDD survey of 424 respondents, we randomly assigned half to receive a set of 11 questions in the forced choice format, while the other half received the same questions in the more traditional, respondent-friendly yes-no format. We recorded the total length of the interview and also asked interviewers to evaluate the ease of administration and respondents' reactions to the survey. These data help determine if there is a significant difference in administration time and administration difficulty between respondents who received the forced choice format versus those who received the yes-no format. In addition, we analyze all 11 questions to document the occurrence of acquiescence, and compare the predictive validity of the two formats. We address the question of whether the improved data quality offered by forced-choice questions warrants the additional time and effort required to administer them.

Grappling with Grids: How Does Question Format Affect Data Quality and Respondent Engagement?

Jene' Grandmont, MarketVision Research (jgrandmont@marketvisionresearch.com); Bob Graff, MarketVision Research (bgraff@marketvisionresearch.com); Lynn Goetzinger, MarketVision Research (lgoetzinger@marketvisionresearch.com); Kara Dorbecker, MarketVision Research (kdorbecker@marketvisionresearch.com)

An ongoing concern over survey quality with focus on online sampling and survey design demands a broad analysis of all components of the process. To identify what matters and to what degree it impacts survey data quality, it is necessary to take all the individual components of this broad perspective and break them down into specific parts taking special care to look within each area for potential quality drivers.

We know online sample can be managed improperly and lead to biased results. We know survey design can be done improperly and skew responses. We know the interview length with online surveys can lead to satisficing or incomplete response, higher drop-out rates, and less satisfaction with the online survey respondent experience. We know there are many components to providing quality and we strive to do

what we can as an industry to effectively determine adequate procedural guidelines as we forge ahead with online survey research. For the purpose of this research presentation proposal we have chosen to present findings related to respondent engagement with particular attention paid to the role of grids in the research design and its impact on quality.

Our research findings are from a large internal R&D project conducted with our proprietary online research panel, Viewpoint Forum. In this research we conduct a multi-cell test to identify alternate ways of presenting attributes and the effect of the alternatives on data quality and respondent engagement. Key findings suggest a specific approach to survey design to maximize both data quality and engagement.

This presentation is focused specifically on this grid research and we hope to contribute to the larger body of knowledge that exists for how to effectively design a research study for an online modality.

Effects of Gender-of-Voice on Reports of Sensitive Behaviors: Findings From the ACASI Module of the Midwest Young Adult Study

Jessica Price, University of Wisconsin Survey Center (jprice@ssc.wisc.edu); Jennifer Dykema, University of Wisconsin Survey Center (jdykema@ssc.wisc.edu); Kerryann DiLoreto, University of Wisconsin Survey Center (kdiloret@ssc.wisc.edu)

While previous research indicates that audio-computer assisted self-interviewing (ACASI) yields higher reports of sensitive behaviors including abortions and sex partners, few studies have examined the potential effect of the gender of the ACASI interviewer's voice on survey reports. Further, we are aware of no research that examines gender-of-voice effects in an ACASI administration among young adults, a population for whom gender might be particularly salient and influential.

In this paper we examine gender-of-voice effects for a set of sensitive questions administered via ACASI as part of the first wave of data collection of the highly successful Midwest Young Adult Study. Respondents include approximately 775 foster care youth (aged 17 at Wave 1 in 2002) who were asked to participate in an in-person study about their experiences while wards of several states in the Midwest. Respondents were randomly assigned to listen to (while simultaneously viewing) approximately 30 questions read by a human female versus a human male voice. (The design crossed respondent's gender with the gender of the ACASI interviewer's voice in order to test for potential interaction effects.) Questions asked about delinquency, violence, and pregnancy.

The analysis explores the effects of the gender-of-voice on several outcomes including survey response distributions, levels of item-missing data, and question completion times. For the analysis of response distributions we examine the effects of a female versus a male voice on both reports of occurrence (the proportion reporting ever engaging in the behavior) and frequency (the mean level of reporting among those who engaged in the behavior more than one time). Contrary to previous research, preliminary analyses indicate some significant voice effects for particularly violent behaviors. Our analysis adds to the relatively small body of research on gender-of-voice effects in surveys using ACASI, with important findings on measuring sensitive behaviors among young adults.

Should Respondents be Offered a Choice, No Choice, Or Their Preferred Survey Mode?

Does Giving People Their Preferred Survey Mode Actually Increase Survey Participation Rates?

Kristen Olson, University of Nebraska-Lincoln (kolson5@unl.edu); Heather Wood, University of Nebraska-Lincoln (hmwood74@yahoo.com); Jolene Smyth, University of Nebraska-Lincoln (jsmyth2@unl.edu)

Survey research has long grappled with the concept of survey mode preference, the idea that a respondent may prefer to participate in a survey in one mode over another. In the last decade, the idea of providing respondents with their preferred mode has gained traction and become increasingly influential in survey design. However, very little research has empirically examined whether providing respondents with their preferred mode actually increases response rates, affects nonresponse bias, or leads to data collection efficiency. This paper experimentally examines the effect of mode preference on response rates and nonresponse bias of survey estimates. Respondents to the 2008 Nebraska Annual Social Indicators Survey (NASIS), conducted by telephone (n=1,811; AAPOR RR3=38%), were asked their mode preference for future survey participation. These respondents were subsequently followed up by two independent survey requests. The first request was a telephone survey, the 2009 NASIS (n=548; AAPOR RR1=54.8%). In the second, the 2009 Quality of Life in a Changing Nebraska (QLCN) survey (n=556; AAPOR RR1=45%), respondents were randomly assigned to one of four mode treatments; web only, mail only, web followed by mail, and mail followed by web. The combination of these surveys permits an empirical test of whether offering respondents a survey in their preferred mode (as the only mode, as the first mode, or as the second mode) increases response rates and affects nonresponse bias on key survey estimates. Preliminary analyses indicate, for example, that those who prefer the web mode participate at twice the rate as others when given only the web questionnaire. We also examine whether offering respondents their preferred mode affects the efficiency of the data collection process in terms of the number of days in the field before completion and the number of follow-up attempts.

Improving Response to Mail and Web Mixed-Mode Surveys: The Effects of Offering Choice on Survey Response

Morgan M. Millar, Washington State University (morgan_millar@wsu.edu); Don A. Dillman, Washington State University (dillman@wsu.edu)

Several general population surveys have shown that offering individuals a choice of responding to a survey via web or mail produces lower response rates than offering mail as the sole response mode. In this paper we present results of two experiments designed to assess the effect of offering a choice between mail and web response when only mail contacts are used, as is necessary in general public populations. The population for these experiments (undergraduate students at Washington State University) was chosen to overcome a limitation of previous studies, i.e. the fact that some members did not have web access. In support of prior theory and research, this research suggests that offering a choice produces a slightly lower response rate than that obtained when only offering a mail response option. At the very least, there is no evidence to support the common assumption that offering a choice will improve response. We also find that offering only a web response option produces a significantly lower response rate than either the mail-only or choice option. Both experiments also include additional treatments that reveal whether the negative effects of choice and offering web-only can be overcome through combining postal contacts, token financial incentives, and strategically timed emails to support web response. Results of these experiments have significant, but different, implications for surveys of general populations in which email contacts cannot be used (mail-only may be a better strategy) and situations in which postal and email contacts can both be used (a web-only response option is likely to be better).

Incorporating a Web Option in a Two-Phase Mail Survey

Pat Dean Brick, Westat (patdeanbrick@westat.com); David Cantor, Westat (davidcantor@westat.com); Daifeng Han, Westat (daifenghan@westat.com); Maribel Aponte, Department of Veterans Affairs (Maribel.Aponte@va.gov)

Industry-wide, RDD surveys have been plagued by declining response. To address this decline, alternative survey designs have been developed. Many of these designs have been fielded without the benefit of precursor methodological studies. This paper reports on the results of development efforts for a two-phase survey conducted for the Department of Veterans Affairs. The first phase of the survey mailed a short screening instrument that identified Veterans in the household. A second survey was sent to those Veterans identified in the screener. The first part of this paper discusses the results of an experiment that varied the mode of the initial request at the screener. One half of the screener sample was asked to complete the questionnaire on the Web, while the other half was sent a paper questionnaire with no reference to the Web. Web non-respondents were sent a paper questionnaire after one reminder notice. Data from the completed pilot demonstrate that the mail condition generated significantly higher response rates among both the general public and among Veteran households. The paper will discuss these results, response rates across different types of households, as well as the effects of a paper insert that was included as a second experimental factor. The second part of this paper describes the outcome of having screener respondents choose the mode for the second-phase survey. At the end of the screening interview, eligible respondents were asked whether they would prefer being mailed a paper survey or would prefer the Web option. We will present data on who chose to use the Web vs. mail, the response rate of those that chose the Web vs. mail and the extent to which the Web respondents eventually completed a paper questionnaire (after extensive follow-up).

Does Providing a Choice of Survey Modes Influence Response?

Virginia M. Lesser, Department of Statistics - Survey Research Center (lesser@science.oregonstate.edu); Lydia Newton, Survey Research Center (newton@science.oregonstate.edu); Danny Yang, Survey Research Center (dyang@science.oregonstate.edu)

Providing a respondent with a choice of modes for submitting a questionnaire would appear to improve response rates and thus reduce nonresponse. A survey of registered recreational boat owners was conducted using different modes and mode choice in 2009. The purpose of this survey was to document the yearly use and gasoline consumption of Oregon registered boats. We randomly assigned three mode groups to investigate whether providing a mode choice had an impact on reducing nonresponse. One group received printed versions of the questionnaire and was asked to return the completed form by mail. This group was not given a choice to complete the questionnaire by any mode other than mail. A letter was mailed to the second group providing them a link to a Web site with the questionnaire. No printed version of the questionnaire was sent with the first mailing. Nonrespondents in this second group were sent a printed version of the questionnaire. The first mailing for the third group offered the option of completing the questionnaire by either Web or by a printed version of the questionnaire. This group received a letter providing them the link to a location of a Web site with the questionnaire and a printed version of the questionnaire. We compared response rates at different stages of data collection, item nonresponse, and responses across the three modes. We also compare the 2009 results with a study of the same population conducted in 2007.

AAPOR Poster Session - 1

Thursday, May 13

3:00 p.m. - 4:00 p.m.

Poster Session - 1

Participation or Communication? Political Activity in the Internet Age

Lindsay H. Hoffman, University of Delaware (lindsayh@udel.edu); Marie Hong, University of Delaware (mhong@udel.edu); Courtney Long, University of Delaware (cml@udel.edu)

Each introduction of a new media technology has shifted public opinion and political behavior in both small and monumental ways, particularly in the context of presidential elections. The 2008 election featured the novel use of a great deal of new technologies. The prevalence of text messaging, social networking and other forms of new media has given rise for the need to reevaluate the traditional definition of political participation. An ever-increasing amount of research on political communication and participation has addressed the rise of new media and its effect on political outcomes, yet much of this research neglects to adequately differentiate communication and participation in the new media environment. Is becoming a fan of a candidate on Facebook participation, or a way to express one's opinion? What about signing an online petition? This research draws from the model developed by Verba, Schlozman, and Brady (1995) to operationalize online participation by analyzing an individual's resources (time, money, and civic skills) as well as networks of recruitment.

We also use theories of computer-mediated communication (e.g., Walther, 2008) and research in political communication (e.g., Blumler & Kavanagh, 1999) to flesh out an operationalization of online activities that can be defined as strictly communication behaviors. Our research utilizes data from a 2008 Pew Internet and American Life project in order to explore the distinctions between these behaviors. We investigate how these new forms of "e-participation" fit into the traditional model of participation, or whether they are better defined as communication activities. We also examine generational differences in these behaviors, as well as overall patterns of behavior. Implications for both measurement and theory are discussed.

Modeling State Differences in Response Rates in RDD Surveys: An Analysis of Seven Years of BRFSS Response Rates

David R Johnson, Pennsylvania State University (drj10@psu.edu); Kurt D Johnson, Survey Research Center (kdj11@psu.edu); Breandan Jennings, The Pennsylvania State University (btj112@psu.edu)

The Behavioral Risk Factors Survey System (BRFSS) is the largest RDD telephone survey conducted in the United States and is designed to produce estimates of health risks at the state and national levels. Because the BRFSS is conducted using common procedures with samples in each state over several years, it provides an excellent sample for modeling state-level differences in response rates. In this paper we use fixed and random effects pooled time-series approaches to model factors that predict state-level response rates over seven years. The response rate data were taken from the BRFSS summary data quality reports available on the CDC website. These contain detailed information on calling dispositions and components of nonresponse for each state. We used data from 2002-2008 because response rates components were reported similarly in these years. Predictors include in the model are the type of organization conducting the survey, state population demographic and economic characteristics taken

from the Census data, political affiliation and voting data, and state response rate data derived from the American Community Survey. State differences in several components of the response rates (e.g. the refusal rate, non-contact rates, and ineligibility rates) were used as outcomes in the analyses to evaluate if state characteristics predicted these in different ways. The data for the random effects models are pooled across the seven years providing more stable estimates of state differences. Fixed effects models were used to estimate the effects of changes in survey provider on components of the response rate. The results will show how well differences between states can be predicted and whether there are consistent state-level differences that cannot be accounted for by these models. The findings will increase understanding about factors predicting response rates and allow researchers planning state level survey to more accurately predict expected response rates.

Investigating Data Falsification: Modes of Detection and Causes of Suspicion

Elizabeth L Love, U.S. Census Bureau (elizabeth.l.love@census.gov)

Data falsification occurs when an interviewer deliberately fabricates any data collected during interviewing or intentionally departs from survey procedures. Research from interviewer management identified common modes of detection that led to definitive findings from investigations of falsification. However, there is scant information available that describes the joint impact of different detection techniques and final results from verifying falsification. In addition, little research has examined the association between causes of suspicion and conclusive findings from formal investigations. The current research primarily concentrates on the combined effect of detection modes on investigative results. Associations between causes of suspicion and findings from formal investigations of falsification will be examined as a secondary component.

Tracing Young Adults for a Telephone/Web Survey

**Benjamin T. Phillips, Steinhardt Social Research Institute, Brandeis University
(bphillips@brandeis.edu)**

Young adults represent a particularly difficult population to follow in longitudinal designs. In this study, a sample of young adults who applied to a program designed to build ethnic/religious identity (Taglit-Birthright Israel) between 2000 and 2003, when they were ages 18-26, were surveyed in February-June 2009 to evaluate program impact. The gap between program application and data collection covered significant transitions in the lives of these young adults, including graduation from college, pursuit of postgraduate education, employment, and marriage, all of which degraded the quality of the contact information collected at time of program registration. To avoid coverage error due to poor contact rates, an extensive set of data enhancement and tracing activities were undertaken. This proposal examines these efforts in detail and considers their effectiveness. The initial step was to standardize and update the registration data via a third-party vendor, using the USPS NCOA and Extended NCOA databases, as well as telephone appends. The first contact attempt was via email to all email addresses supplied at time of registration. Telephone contacts then began with to telephone numbers supplied at time of registration and the appended telephone numbers. Where a family member or other person with knowledge of the target individual was located, new contact information was sought. In other cases, additional steps to collect data were undertaken. Two "people search" services were used, Intelius and PublicRecordsPro, together with an email address database (emailfinder.com), regular white pages listings, social networking sites (Facebook and LinkedIn), as well as general Internet searches for the individual. A final contact rate (CON2) of 62 percent was achieved, much higher than would have been the case without extensive follow-up.

Correlation of Roster Probes and Population Complexity

Celeste Stone, U.S. Census Bureau, Decennial Statistical Studies Division
(celeste.nicol.kline.stone@census.gov); **Kyra Linse, U.S. Census Bureau**
(kyra.m.linse@census.gov)

In 2009, U.S. Census Bureau staff field-tested a post-enumeration survey instrument that will be used to measure the coverage of the final 2010 decennial census counts. One primary goal of the interview is to collect names of all current residents, particularly those who might have been missed during the census. Drawing on previous research indicating that certain groups in the population are more likely to be deliberately (Tourangeau et al., 1997) or accidentally (Martin, 1999) omitted from household rosters, the interview uses methodological innovations like question probes to help respondents identify a complete and robust roster of all residents.

This paper discusses the historical development of those probes and evaluates their current utility by evaluating key attributes (e.g., tenure, sex, age, race, and residence status) of the people collected using the roster probes used in the 2009 Person Interview Operational Test. We assess the extent to which the probes capture the subpopulations typically underreported when standard survey procedures are used. Of special interest for this test is the quality of the residence information collected for persons captured by the probes (e.g., move dates, other “usual residences”, explanations of complex living situations, etc.) needed to produce good coverage estimates. Finally, to the extent possible, reinterview data is used to assess consistency of reporting of tenuously attached people over multiple measurements. Findings provide timely updates to the methodological research at a time when economic factors might be increasing the numbers of tenuously attached persons throughout the U.S.

Impact of Interviewer and Respondent Verbal Behaviors on Data Quality in Different Interviewing Techniques

Ipek Bilgen, University of Nebraska-Lincoln (bilgenipek@yahoo.com); Robert F. Belli, University of Nebraska-Lincoln (rbelli2@unlnotes.unl.edu)

A nationwide subsample of 626 participants was randomly selected from the 2001 Panel Study of Income Dynamics (PSID). These participants were interviewed via telephone retrospectively about their life course events by either Event History Calendar (EHC) or Standardized Conventional Questionnaire (CQ) methods. Among these interviews randomly selected 165 EHC and 162 CQ interviews have been verbal behavior coded. This paper compares the use of interviewer and respondent behaviors in different interviewing techniques and how this might impact the data quality. Data quality is measured by validating the experimental retrospective reports against reports in the traditional PSID interviews provided annually on the same topics by the same respondents. In this study, we hypothesize two separate factor structures for interviewer and respondent verbal behaviors. The factor structure for interviewer behaviors contains three constructs: 1-Retrieval probes (which include parallel, sequential and duration strategies); 2-Conversational Probes (such as interviewer verification, clarification and task related feedback strategies which aim to satisfy questionnaire objectives); 3-Rapport behaviors (such as interviewer laughter, digressions and feedbacks which do not necessarily aim to satisfy the questionnaire objectives). The factor structure for respondent behaviors also consist of three constructs: 1-Interviewing Difficulty Behaviors (which include respondent request for clarification, corrections, disagreements, and don't knows); 2-Retrieval Strategies (include probes such as respondent parallel, duration, sequential and respondent use of age and season information) and; 3-Rapport Behaviors (such as respondent laughter and digressions). Preliminary findings reveal that the hypothesized interviewer and respondent behavior factor structures have an acceptable fit in both EHC and CQ interviews. Moreover, the preliminary results also illustrates that both interviewer and respondent retrieval strategies significantly increase the data

quality in EHC interviews; whereas, these behaviors have no significant impact on data quality in CQ interviews.

Creating a Self-Administered Event History Calendar

Megan Zander Cotugno, RAND Corporation (mzander@rand.org)

Information from respondents' pasts was critical to our recent research project. The goal was to understand the reasons individuals enlist in the Army. Of particular interest were "older" recruits: those who could have enlisted at age 18 but instead enlisted later. In addition to responses to survey questions, we wanted to collect detailed academic, social, and economic information from the time participants left high school to the time they enlisted in the Army to understand what happened in these individuals' lives that prompted them to enlist later in life. With Army guidelines allowing enlistment up to age 42, this could be 25 years of life history for some participants to recall.

Event history calendars (EHCs) aid respondents' memories by giving visual cues that help them place past events on a timeline. Most literature on EHCs focuses on interviewer-administered tools, but cost and time constraints prevented us from using a one-on-one, interviewer-administered model for collecting EHC data. So we designed a version that could be self-administered in group sessions.

Through numerous iterations, multiple pre-tests, and extensive use of visual aids, we modified a questionnaire that is typically interviewer-administered into a questionnaire that could be completed as a self-administered survey with proctor guidance during a group survey session. I will detail the major differences in administration and instrument format that resulted from our use of a self-administered EHC rather than an interviewer-administered EHC and will discuss the quality of data collection with our modified EHC, some important lessons learned, and the components of a successful self-administered EHC.

Parent Proxy and Adolescent Self-Report: Parental Judgment of Adolescent Knowledge and Perceptions of Disease, Treatment, and Medical Care

Laurie Fishman, Harvard Medical School/Children's Hospital Boston

(laurie.fishman@childrens.harvard.edu); Sonja Ziniel, Harvard Medical School/Children's Hospital

(sonja.ziniel@childrens.harvard.edu); Sarah Krathwohl, Children's Hospital Boston

(sarah.krathwohl@childrens.harvard.edu); Janis Arnold, Children's Hospital Boston

(janis.arnold@childrens.harvard.edu)

Health surveys that examine health and disease related behaviors historically have used parent proxy-report in the stead of adolescent self-report. Indeed, in certain situations, parents provide accurate health information, while under other circumstances adolescents have been shown to report accurately on their health and sometimes more so than a parent. There is, however, little research on the accuracy of a parent's judgment of their adolescent's knowledge about their disease and treatment process. It is therefore unclear when survey researchers should use adolescents to report their health behaviors or when parents should be used as proxy reporters. In a study conducted at the Harvard Medical School/Children's Hospital Boston a survey of adolescents aged 10 and older with inflammatory bowel disease and their parents was conducted to assess parent and adolescent knowledge of the disease, treatment, and medical care, and parent's judgment of their adolescent's knowledge and perceptions about these. A sample of 250 adolescents and parent dyads completed self-administered questionnaires including objective questions about the adolescent's condition, treatment, and symptoms, as well as both objective and subjective questions about adolescent involvement in self-care and physician visits. A subset of questions from the adolescent and parent surveys was compared to hospital electronic medical

records as the “gold-standard” to confirm diagnosis, treatment, and symptoms related to the disease. As previous literature indicates, there is generally good agreement between adolescent and parent responses. However, in examining parental judgment of adolescent knowledge compared to electronic medical records, parent reports differ from the adolescent’s knowledge of his or her disease and treatment process. The relationship of demographic factors with respect to accurate reports from parent and adolescent is also assessed. The results of this study may provide insights into the utility of adolescent self-report and parent proxy-report in health surveys given different circumstances.

A Comparison of Web and Mail Survey Respondents Within a Mixed Mode National Survey
Eunjung Shin, UIC (eshin9@uic.edu); Timothy P. Johnson, UIC (timj@uic.edu); Kumar Rao, Gallup (kumar_rao@gallup.com)

Because of their cost/time-efficiencies, web surveys are being increasingly incorporated into national survey data collection programs in the United States. Yet, coverage and data quality issues in web surveys remain important challenges. As a basic study designed to better understand data quality in a mixed mode national survey, this paper investigates the (1) socio-demographic characteristics of web vs. mail respondents, and (2) the effects of survey mode on response rates and data quality in a 2008 Gallup Health panel survey. The Gallup panel examined is selected with probability-based sampling from the U.S. general population and employed to conduct a mixed mode web-mail survey (number of respondents = 39,444). Whether panelists complete a web or mail questionnaire is pre-assigned according to their internet-use patterns, which were identified at the time of recruitment.

Findings indicate that the mode selection procedure produces significant socio-demographic differences between web and mail respondents. Web respondents’ characteristics were found to be significantly different from those of mail respondents in terms of age, gender, race, household-income, education level, urbanity and the duration of panel participation. After multivariate adjustments for these differences in sample composition, mode effects on response rates continue to be documented. Although the web survey mode produced significantly lower response rates overall, compared to the mail survey mode, respondents in the youngest age group were nonetheless more likely to respond when assigned to the web survey mode. In addition, when examining data quality via open-ended response patterns, some evidence was found indicating that the web survey mode may produce higher quality information, although multivariate controls suggest these mode effects may be limited. Our presentation will include consideration of future research needed to improve the quality of mixed mode surveys that include a web-based component.

Question Wording and Support for a Health Insurance Public Option
Alan Reifman, Texas Tech University (alan.reifman@ttu.edu)

The domestic centerpiece of President Barack Obama’s first year, reflecting a Democratic Party goal since 1945, has been pursuit of healthcare reform achieving universal coverage. Some congressional proposals have sought a new “public option” (PO) program to complement existing public and private insurance plans. The PO has been particularly contentious in Congress, with the furor implicating public-opinion polling as PO proponents and opponents have debated whether question wordings have inflated or depressed popular support for the idea. The present study used multiple-regression techniques with poll as the unit of analysis, to see if the presence or absence of certain phrases in survey items were associated with PO support, controlling for other poll aspects (pollster identity, dates in the field, and adult vs. voter sample). Specifically, analyses examined whether items: presented the PO as providing a choice to consumers; stated that interested consumers would have to pay to join; likened the PO to Medicare; and described the PO as competing with private insurance. Forty-two polls from July-October

2009 were included, with pollsters who surveyed infrequently on the PO and had outlyingly high or low support values excluded (the Durbin-Watson test was used to account for possible autocorrelation). The intercept was 49.3, indicating the average support for the public option with all predictors held at zero (the “regular” mean was 54.2). Four predictors were significant ($p < .005$), with their unstandardized coefficients showing how the presence of a given feature was associated with change in percent support, relative to the intercept: NBC/Wall St. Journal polls ($B = -6.73$), Economist/YouGov polls ($B = -12.80$), items describing the PO as a choice ($B = +6.67$), and items likening the PO to Medicare ($B = +6.58$). The low support in Economist/YouGov polls should be interpreted in light of their having high (20-30%) undecided percentages.

Exploring Race and Hispanic Origin Questions With Multiracial Individuals: Diverse Responses From a Diverse Population

Rodney L Terry, U.S. Census Bureau (rodney.terry@census.gov); Jennifer Hunter Childs, U.S. Census Bureau (jennifer.hunter.childs@census.gov)

The U.S. Census Bureau continuously works to develop race and Hispanic origin questions that are applicable to respondents at the time of administration. This involves designing questions that attempt to reflect the evolving complexity of race and Hispanic origin identification. This is particularly important for a growing multiracial population that is estimated to comprise twenty percent of the U.S. population by 2050 (Smith & Edmonston, 1997). Researchers within sociology, demography, and psychology are beginning to examine the fluidity of multiracial individuals’ responses to race and Hispanic origin questions, while hypothesizing several factors that may result in such fluidity. These factors include having more categories to choose from, the effect of previous race-based social interaction on how multiracial individuals respond to these questions, and their perceptions of how they are perceived by others. Cognitively testing these types of questions may contribute important knowledge about how these factors present themselves in questions designed for the census.

Census Bureau staff developed and cognitively tested experimental versions of race and Hispanic origin questions that will be used during the 2010 Census, as well as a telephone reinterview that will be used to evaluate the experimental questions. These experimental questions allow respondents to self-identify race in response to alternative question wording that may be more aligned with contemporary views of race and Hispanic origin. The telephone reinterview offers several opportunities for each respondent to report his or her own race, as well as how his or her race or origin is perceived by others. The cognitive interview debriefings from both projects focused on self-perceptions of race and Hispanic origin and how it might differ in different contexts. This paper summarizes cognitive testing results from both projects with 20 multiracial respondents. Key findings discussed include response variance when compared across different interview and debriefing questions.

¿Habla Español? Impact of Offering a Bilingual Option in a Mail Survey of Linguistically Isolated Areas

Andrew Zuckerberg, National Center for Education Statistics (andrew.zuckerberg@ed.gov); Daifeng Han, Westat (DaifengHan@westat.com)

The National Household Education Survey (NHES) is undergoing a conversion from a random digit dial telephone survey to an address based mail survey with telephone non response follow up. The survey requires screening sampled households to determine the presence of eligible children. If eligible children are present, within household sampling is performed to select a reference child. On the telephone, computer assisted interviewing allows a seamless transition between screening and interviewing. However, the conversion to a mail survey requires separating the screening and interviewing process into

two phases. The screening is performed in a brief initial questionnaire. Households with eligible children are then sent a follow up topical interview. In the 2007 NHES telephone administration, 4.8% of screener interviews were conducted in Spanish. As a result of the change in mode from telephone to primarily mail administration, it was thought that a Spanish language option would be necessary for the interviews. This paper will look at the results of an experiment that compared a bilingual (English and Spanish) screener questionnaire to an English only form in a special sample of addresses in census tracts with a high density of linguistically isolated (Spanish speaking) households. Respondents were randomly assigned to one of the two treatment groups. Participants in both groups received a reminder post card one week after the initial questionnaire. The postcard gave respondents the opportunity to request a Spanish version of the questionnaire. In addition, respondents could call a toll free number to request a Spanish language version. This paper will present the results of offering both options, information on which language respondents completed the bilingual form in, as well as differences in demographic and household characteristics of respondents to the two forms.

“¿Podemos Comenzar La Entrevista Ahora? / May We Begin The Interview Now?” Conducting Survey Research With Bilingual Interviewers and Respondents

Nicole Tate, RTI International (ntate@rti.org); Natasha Janson, RTI International (njanson@rti.org)

Conducting surveys with Spanish-speaking populations can be challenging in terms of instrument design as well as data collection and interviewer training and support. For a 2009 large-scale longitudinal education study of approximately 17,000 recent bachelor's degree recipients, an abbreviated Spanish interview was offered as an option to sample members. The study accommodated Spanish-speaking sample members by providing a Spanish language web interview, as well as telephone and field interviews.

This poster demonstrates the challenges involved in creating a methodologically-equivalent non-English language interview.

Questions include:

- Is the integrity of question concepts maintained post-translation?
- How will dialect differences be accounted for post-translation?
- How are non-translated interview items accounted for within the instrument design?

We also discuss challenges met in training and supporting bilingual interviewing staff. Questions include:

- What special recruiting techniques are necessary to hire competent bilingual staff?
- How much supervision is necessary to accommodate bilingual interviewers?
- Is it necessary to translate all study materials for bilingual interviewers?
- To what degree should interviewers be expected to translate “on the fly?”
- Are varied refusal conversion techniques necessary when contacting a bilingual population?
- What cultural factors must be taken into account?
- What special considerations are needed when supporting bilingual interviewers?

Based on successes and challenges with this Spanish interview, we will also provide suggestions and topics for discussion for other multi-mode, multi-lingual interviews.

No, No Cash Rent

Julie Parker, U.S. Census Bureau (julie.parker@census.gov)

According to the American FactFinder, almost 6% of the US renter-occupied housing units claimed that they did not pay rent. This paper examines the American Community Survey (ACS) tenure question and

uses matched administrative data from the Department of Housing and Urban Development (HUD) to investigate whether some households responded incorrectly to the ACS while receiving HUD subsidies. This study focuses on the non-rent-paying population and looks at demographic characteristics, socioeconomic conditions, and living arrangements.

It is important to evaluate the accuracy of responses to the tenure question, as incorrect responses will bias housing cost estimates for renters. The housing cost-to-income ratio (HCIR) is calculated as the gross rent divided by monthly household income. This statistic is not calculated for households with “no cash rent” (NCR), even if they have utility and fuel expenses. The HCIR is a way to estimate the level of housing cost burden; excluding households that responded incorrectly would underestimate this burden and produce inaccurate estimates of housing affordability (Wendrip and Pelletierre, 2009).

The research is conducted using probabilistic record linkage techniques on 2006-2008 HUD administrative data and the 2005-2007 ACS. The HUD data files used were Public and Indian Housing Information Center (PIC) and Tenant Rental Assistance Certification System (TRACS). Preliminary findings show that of the 103,000 HUD households in the pooled sample, less than 3% claimed they did not pay cash rent. This small percentage suggests that HUD recipients are correctly answering the tenure and rent paid questions.

Racial and Ethnic Differences in On-Schedule Mammography Screening: Results From the Behavioral Risk Factor Surveillance System

Michelle L Rogers, Brown University (michelle_rogers@brown.edu); Melissa A Clark, Brown University (melissa_clark@brown.edu); William Rakowski, Brown University (william_rakowski@brown.edu); Crystal D Linkletter, Brown University (crystal_linkletter@brown.edu)

Several studies show that Hispanic and Black women have higher adjusted odds ratios for mammography screening compared to Non-Hispanic White women, even though their crude screening rates are typically lower than those for White women. This study sought to identify factors associated with this reversal in association, examining the effects of county-level poverty and an individual's access to health care. We used the 2006 and 2008 Behavioral Risk Factor Surveillance System (BRFSS) Selected Metropolitan/Micropolitan Area Risk Trends (SMART) data, which provides information on individuals residing in selected MMSAs with 500 or more respondents and allows analysis at the county-level. Counties with 20% or more of the population identified as Black or as Hispanic were classified as high Black and high Hispanic, respectively. We compared high Black and high Hispanic counties to White counties. We found no relationship between county-level poverty and the likelihood of on-schedule mammography screening in White counties, nor in high Hispanic counties. We did find, however, that women living in high Black counties were significantly less likely to be on-schedule if that county had a high level of poverty. Not surprisingly, lacking health insurance coverage (HI) or a usual source of care (USOC) was associated with a significantly decreased likelihood of being on-schedule. However, in White counties, we found that minority women without HI or a USOC were more likely to be on-schedule than were White women without HI or a USOC. This relationship did not hold, however, in high Black or high Hispanic counties. We argue that this relationship is driven largely by the efforts of programs such as the National Breast and Cervical Cancer Early Detection Program, which target minority women and those without HI. These targeted efforts may be easier to implement in predominately White counties due to less women needing the assistance.

Polling the Poll Workers: Examining the Role of Election Day Workers in 2008

Stephen T. Mockabee, University of Cincinnati (stephen.mockabee@uc.edu); Quin Monson, Brigham Young University (Quin.Monson@byu.edu); Kelly Patterson, Brigham Young University (Kelly_Patterson@byu.edu); Lisa Argyle, University of California, Santa Barbara (lisa.argyle@gmail.com)

The Help America Vote Act standardized voting equipment and registration systems, but did little to address poll worker recruitment, training, or retention. However, poll workers are critical because they interpret the law, affecting what kinds of identification are accepted and who casts a provisional ballot. In addition, prior research finds that poll workers significantly affect the quality of the voting experience as reported by voters. Despite growing demands on poll workers, most jurisdictions have not updated their training methods. With support from the Pew/JEHT Make Voting Work initiative, the authors studied election administration in counties in Ohio and Texas in 2008 to determine the extent to which various types of training can improve poll worker performance. As part of these studies, poll workers were surveyed following the election using a mixed mode approach (Web and US mail). Drawing on survey data from nearly 5000 poll workers, we examine demographic characteristics, attitudes toward politics and government, motivations for working at the polls, evaluations of and responses to training, satisfaction with their jobs as poll workers, and confidence in their own ability to carry out their duties. To supplement the survey data, we draw on data from post-training quizzes on factual knowledge, Election Day evaluations of poll-worker performance in precincts, and post-election focus groups conducted with poll workers. In an experimental design, poll workers were randomly assigned to complete on-line training in addition to standard classroom training. We find that participating in on-line training increases poll workers' confidence in the electoral process, confidence in their own ability to carry out their duties, and factual knowledge of election procedures. From survey and focus group responses we identify factors likely to increase the recruitment and retention of poll workers in future elections.

Evaluating the Effectiveness of Using an Additional Mailing Piece in the American Community Survey

Jennifer Guarino Tancreto, U.S. Census Bureau (jennifer.guarino.tancreto@census.gov)

Decreases in response to mail surveys over the years coupled with the costliness of conducting nonresponse follow-up personal interviews has created a need for rethinking the way we have traditionally conducted mail surveys. Despite being a mandatory survey, the American Community Survey (ACS) has not been immune to the recent downturn in mail response rates. As a result, the Census Bureau has identified a need to test changes to the ACS mailing strategy that would potentially increase the ACS mail response rate and reduce the costs of the Computer-Assisted Personal Interviewing (CAPI) nonresponse follow-up operation. The 2009 ACS additional mailing test addressed this need by determining the feasibility of incorporating an additional reminder postcard versus a third mailed questionnaire to boost the mail response rate for nonrespondents without telephone numbers. Using the March panel from the 2009 production ACS sample, the additional mailing test universe included any cases for which we had no telephone number data and that were identified as a mail nonresponse approximately two weeks after the mailout of the current replacement questionnaire. Through the additional mailing test, we demonstrated that by adding either an additional replacement questionnaire or an additional reminder postcard we can increase the ACS mail response rate for households without telephone numbers by about 6 to 7 percentage points. This paper describes the test treatments, experimental design, and results of this test.

Relationships Among Social Science Teachers' Demographic Characteristics, Demographic Characteristics of Their Respective Schools, and Teachers' Opinions and Behaviors as Related to Field Trips: A Case of Ohio

Anastasia Mirzoyants, The University of Toledo Urban Affairs Center
(anastasia.mirzoyants@rockets.utoledo.edu); **Dwight Haase, The University of Toledo Urban Affairs Center** (dwight.haase@utoledo.edu)

This study investigated the general trends in the opinions and behaviors of social science teachers in the Ohio K-12 system with regards to field trips. The additional goals of the research were to define whether there were relationships among those opinions and behaviors and three groups of demographic characteristics: personal characteristics of the teachers, their professional characteristics, and characteristics of their respective schools. Among these characteristics, the author looked at such attributes as teacher age and gender, educational attainment and professional experience, and the school-specific percentages of minority students and students receiving free/reduced lunches. In addition, the four types of school support for field trips were examined for their relationships with the teachers' opinions and behaviors.

The data collected through a web-administered survey was analyzed using SPSS and WinSteps software. The analyses of the data supported all six study hypotheses; in other words, the researcher found statistically significant relationships between individual characteristics within three groups of demographics and the teachers' opinions and behaviors related to field trips. The group of factors that correlated the most with the mentioned opinions and behaviors included the teachers' age, gender, professional experience, the schools' type, and the percentages of minority students and students receiving free/reduced lunches. In addition, the presence and the type of school support activities were found to be related to the teachers' definition of educational value of field trips. The implications of the findings for continuing teacher education and for relationships between the K-12 system and the network of field trip hosts are discussed in the last part of this paper.

Concurring Opinions: An Examination of Differences Among Believers in Climate Change
Christopher Paul Borick, Muhlenberg College Institute of Public Opinion
(cborick@muhlenberg.edu); **Barry G. Rabe, University of Michigan** (brabe@umich.edu)

Do you believe in global warming? The answer to this simple question has sparked large amounts of research and significant levels of debate among policymakers and scholars around the world. Public opinion surveys in the United States have found fluctuating levels of belief in climate change over the past decade, with a period of steadily increasing belief followed by a recent decline in the percentage of Americans accepting this phenomenon. While considerable scholarly and media attention has focused on determining who believes and who doesn't believe in global warming, there has been far less consideration given to the exploration of differences among the populations of believers in climate change. Yet at any given time the group of Americans that believes in global warming is made up of individuals that share a common belief in the outcome (i.e. a warmer planet), but differ on the cause of the warming that they believe is occurring. Among the community of believers in global warming are those that place humans as the sole reason for the changing climate, those that see increasing temperatures as part of a natural cycle, and those that see both natural and anthropogenic determinants of a hotter planet.

In this study we examine differences among believers in climate change through an analysis of the National Surveys on American Public Opinion on Climate Change and Policy Options. First we examine the effect of various individual level characteristics on determining the types of belief in climate change (e.g. human induced, natural causes). Next we explore the impact of various factors on determining beliefs in global warming with an emphasis on isolating the effects of selected stimuli. Finally we turn our attention to the effect that differences in beliefs about the causes of global warming may have on policy preferences on this matter.

Effects of Personalization, Length and Choice of Mode on Cooperation Rates for Mixed Mode Recontact Surveys

Darby Steiger, The Gallup Organization (darby_miller_steiger@gallup.com); Manas Chattopadhyay, Gallup (manas_chattopadhyay@gallup.com); Dawn Royal, Gallup (dawn_royal@gallup.com)

Since 2008, Gallup has been conducting 1,000 surveys daily using a dual-frame RDD and cell phone sampling. With an annual sample of approximately 355,000 respondents, this research undertaking affords a great opportunity to tap into an already-cooperative, nationally representative sample for additional research studies. Every survey respondent is asked if they would be willing to be recontacted “to learn a bit more about their thoughts and opinions.” Typically, about 50% agree to be recontacted.

In the fall of 2009, Gallup was asked to leverage this recontact sample to conduct a follow-up mail/web survey regarding international travel. An experiment was designed to test various methods to maximize willingness to be recontacted and to provide a mailing address or email address, as well as to boost actual response rates to the follow-up survey. Over the course of 38 days of data collection, an experiment was conducted with three treatments. Each treatment had three conditions to which all daily survey respondents were randomly assigned. The first treatment tested various tailored messages to seek the respondent’s initial agreement to be recontacted based on their responses to substantive items in the telephone survey (such as “We are specifically hoping to follow-up with people like you who are...”). The second treatment tested different messages regarding the average time to complete the follow up study. The final treatment sought to determine the most effective sequence of questions to assign the respondent to either the mail or Web mode in order to maximize the number of valid email and/or mailing addresses provided.

The findings will help improve the size and performance of recontact sample from a large, nationally representative telephone sample.

Sampling by Professional Referral: Lessons Learned From Asking Physicians to Recruit Patients in Two Health Surveys

Sean O Hogan, RTI-International (hogan@rti.org); Michael Power, American College of Obstetricians & Gynecologists (MPower@acog.org); Jay Schulkin, American College of Obstetricians & Gynecologists (jschulkin@acog.org); John D Loft, RTI International (jloft@rti.org)

When there is no reliable list from which to draw a statistical survey sample, the best alternative may be to rely on sampling by referrals. This method is sometimes called snowball sampling (Welch, 1975;), respondent-driven sampling (Heckathorn 2002), or multiplicity sampling (Rothbart, Fine & Sudman, 1982). While much of the literature on referral sampling describes peer referrals, we explore two cases where we relied on professional intermediaries to reach subjects who are part of a rare population.

We compare our experiences conducting two surveys that depended on physicians who were asked to identify patients with specific health conditions and ask the patients to complete a questionnaire. While the methods used in these projects had many similarities, they varied in that one required quota sampling, it involved a relatively rare disease and the length of the questionnaire was somewhat longer. Meanwhile, the other procedure relied on a much more streamlined approach. In the latter case we easily satisfied final sample size requirements, while in the former we encountered much more difficulty. Several related lessons are reported and presented for further hypothesis testing. These lessons learned may be useful for researchers who depend on professional intermediaries to conduct interviews hard-to-reach subjects or subjects with rare conditions.

Google Public Data: A Tool to Search for and Visualize Official Statistics

Mario Callegaro, Google (callegaro@google.com); Ola Rosling, Google (olar@google.com); Jurgen Schwarzler, Google (jurgen@google.com)

Google public data was launched in April 2009 to make public data easier to find and more visible to the user. When somebody is searching for "unemployment rate California", for example, Google will show a special public data search feature above the first result, which - upon clicking - shows a landing page where the user can visualize trend data since 1990, allowing more states and counties to be added for comparison. A link to the current visualization can be shared among users. Initial launch had US population and unemployment rate data, and more recently World Bank indicators for countries and economies worldwide have been added. For each indicator/metric the user can go directly to the provider's webpage and get all the necessary methodological information about the sources.

The goal of this paper is to present the tool to AAPOR community and collect feedback. Specifically the paper will be tailored to use "public data" for benchmarking statistics, possibly weighting, and for educational purposes. For example the tool can be used with students to discuss trends and write papers.

We also want to identify publicly available, well known quality data, irrespective of topic and locale. We are interested in both aggregated statistics and the underlying raw data from which they were derived. Other types of structured information like reference lists and classifications are also of great interest. We will not use any data that compromises the privacy of individuals or infringes upon any proprietary rights. Collaboration with official statistical institutes and providers will increase the number of featured datasets. Educational, not for profit and commercial data publishers are welcomed in collaborating in this effort.

Mixed Mode Data Collection Methods in the American Community Survey

Deborah Harner Griffin, U.S. Census Bureau (deborah.h.griffin@census.gov); Todd R Hughes, U.S. Census Bureau (todd.r.hughes@census.gov)

The American Community Survey (ACS) is a survey of the population living in the United States and Puerto Rico. Demographic, social, economic, and housing characteristics are included in this mandatory survey that is conducted by the Census Bureau. Households are interviewed throughout the year using three sequential modes of data collection. The mixed mode design used in the ACS starts with mail and follows up nonrespondents to the mail by telephone and eventually, in person. This paper describes the effectiveness of the ACS in each of these modes and how the modes work together to achieve high levels of overall survey response. The paper includes mode level analysis of response rates, item nonresponse rates, costs, and workloads. The diversity of the population of the United States and Puerto Rico with respect to language, living situations, and literacy means that a single data collection method is insufficient. Diversity in housing and access to landline phones also means that methods well suited for some households, will not be effective for others. ACS respondent characteristics therefore vary by mode. This paper summarizes what we have learned about the fit of modes to respondents. In addition, an overview of the management of work across each mode is detailed, as are the challenges inherent in mixed mode designs. The Census Bureau is planning to test an ACS web data collection mode to determine if this option could reduce survey costs and improve survey quality. The paper discusses that research and some of the known issues the ACS is facing with adding a fourth data collection mode.

New Developments in Privacy for Online Research

Peter Milla, Independent Consultant (pmilla@optonline.net)

Privacy and respondent confidentiality are core concepts in market and survey research and have historically been addressed in industry organization codes and guidelines. Recently however, with the changes in research (and commerce in general) brought by the Internet, privacy legislation and regulation has made privacy an area of increasing complexity, importance and risk for research companies

Privacy regulation like the EU Data Directive, PIPEDA and the quickly evolving area of Digital Fingerprinting and related activities, present urgent compliance requirements for research companies.

The presentation will examine a series of topics, including:

- The issues impacting privacy
- Discussion of what companies should do to evaluate the need and then achieve and maintain compliance
- Detailed discussion of the area of Respondent Identification Technologies, including Digital Fingerprinting:
 - Discussion of the related privacy and data protection concerns
 - Recommendations for the legitimate application of these technologies
 - Discussion of recent industry research regarding device identification technology

Exploring Visual Effects in Mail Survey Materials on the Quality of Survey Data

Teresa (Ye) Jin, The Nielsen Company (teresa.jin@nielsen.com); Lorelle Vanno, The Nielsen Company (Lorelle.Vanno@Nielsen.com); Jennie Lai, The Nielsen Company (Jennie.Lai@Nielsen.com); Michael W. Link, The Nielsen Company (Michael.Link@Nielsen.com); Charles D. Shuttles, The Nielsen Company (Chuck.Shuttles@Nielsen.com)

The importance of graphic layout, visual elements and format of instructions for survey instruments on the quality of survey data is well established for self-administrated survey instruments. Surveys that are kept for an extended period of time (e.g., a one week radio diary) may see respondent compliance drop off over the course of the week and diminished data quality. Therefore, communicating the objectives and process to respondents plays an especially critical role on the quality of collected data.

The primary goal of this research is to discover whether new mail material design (with improved visual appeal) can help generate more returned surveys and improve data quality. Further analysis focused on discovering whether the respondents with different demographic characteristics react to the new approach differently. The findings of this research will help shed light on how survey researchers can build graphic cues and visual designs into the format of the survey instrument and supporting materials, in order to convey more rich information to respondents and improve the accuracy of the data collected.

This research focused on the survey instrument and supporting materials for Nielsen's radio audience measurement. In an effort to motivate respondents to read and follow instructions properly, reduce unidentified data and improve data quality, the instructions were revised. In the survey instrument, visual elements were added to facilitate the text illustration, unnecessary information was removed from the illustrated examples, and data recording sections were clarified more visibly. In addition, a reference guide was re-designed with images and text illustrations for each key step of the survey process rather than the text-based instructions in the original version. Data collected from two radio measurement markets will be analyzed to compare if the data quality is improved in the new version of the survey instrument and supporting materials versus old versions.

Exploring Visual Effects in Mail Survey Materials on the Quality of Survey Data

Emily McFarlane Geisen, RTI International (egeisen@rti.org); Joseph Murphy, RTI International (jmurphy@rti.org); Jessica Severance, RTI International (jseverance@rti.org); Murrey G Olmsted, RTI International (molmsted@rti.org)

Researchers at RTI International conducted several experiments within a national survey of board certified physicians as part of the 2010 “America’s Best Hospitals” rankings for U.S. News & World Report. This paper discusses the results of two methodological experiments designed to increase physician cooperation and participation.

The first methodological experiment tested the effect of a custom “America’s Best Hospitals” photo stamp on response rates. In the previous year’s survey, the authors found that oncologists who received breast cancer research stamps on either the outgoing or return envelope were more likely to respond compared to oncologists who received regular postage stamps on both the outgoing and return envelopes. Based on those earlier results, it is hypothesized that the custom stamp will be more appealing and noticeable to physicians, thus increasing response rates.

The second methodological experiment tested the use of incorporating a photographic image of a hospital patient on the survey form itself. It is hypothesized that the photograph may encourage participation by providing physicians with an image of a potential patient, and thus viewing the survey as a request for their expert medical opinion on which hospitals provide the best care for their patients.

Mandatory Reporting of Child Maltreatment: A Technique for Risk Assessment and Data Review

Stacey Weger, RTI International (sweger@rti.org); Amy Hendershott, RTI International (hendershott@rti.org); Gina Kilpatrick, RTI International (ginak@rti.org)

In surveys where questionnaire responses may indicate potential child abuse or neglect, it is imperative that the project team develop guidelines and processes to determine whether a report should be made to appropriate authorities. It is of particular importance that the decision to report or not report is weighed against the risk that could be created by an erroneous report. To address and mitigate the risk of filing an erroneous report of abuse or neglect, we developed a mandatory reporting protocol that minimizes the likelihood of such an occurrence. The protocol takes into account the obligation to protect the confidentiality of participants and the obligation to report cases that meet the threshold set for reporting a possible case of abuse or neglect.

The aim of this methodological brief is to outline the procedures used in a study where parents and minor children were asked questions regarding victimization and perpetration of physical and sexual abuse, as well as neglect. In a recent longitudinal study conducted by RTI International, such data were collected, and the study team utilized a unique, multifaceted approach to determine if situations described by respondents – both parents and minor children – were indicative of offenses that may require a formal report to authorities. Automated email reports drawing upon respondent data were used in conjunction with a “decision tree” that defined the parameters for survey item responses to determine whether a case review was required. However, understanding the nuances of individual cases cannot always be achieved by relying wholly on programming; a human element was also integrated in this process in the form of group review and discussion of potentially reportable cases. This methodological brief will describe the mandatory reporting protocol in an effort to share the knowledge and lessons learned throughout data collection on a complex project.

Self vs. Spouse's Report of Alcohol Consumption: An Evaluation of Proxy Reports of Spousal Drinking Behavior

Young Ik Cho, Survey Research Laboratory, University of Illinois at Chicago (youngcho@uic.edu); Richard W Wilsnack, University of North Dakota, School of Medicine & Health Sciences (rwilsnac@medicine.nodak.edu); Sharon C Wilsnack, University of North Dakota, School of Medicine & Health Sciences (swilsnac@medicine.nodak.edu); Arlinda F Kristjanson, University of North Dakota, School of Medicine & Health Sciences (akristja@medicine.nodak.edu); Perry Benson, University of North Dakota, School of Medicine & Health Sciences (pbenson@medicine.nodak.edu); Timothy P Johnson, Survey Research Laboratory, University of Illinois at Chicago (timj@uic.edu)

Self-report has been the primary (if not the sole) source of data regarding individual alcohol consumption patterns. Collateral reports from significant other persons have been used as a valuable source of validation in the absence of a gold standard measure of past drinking behavior. Despite some inconsistent findings, self-reports of drinking behavior in general is highly correlated with collateral reports, and this implies that collateral reports may be valid and reliable proxy reports of drinking behavior. However, not many studies have focused on factors that are related to concordance between self vs. collateral reports. Using a sample of 369 married or cohabiting heterosexual couples in the US, we (1) examine concordance between couples' self vs. spouse's reports of various aspects of drinking behavior, including dichotomous measures of abstinence, as well as continuous measures of frequency and amount of alcohol consumption, (2) compare qualities of spousal reports as proxy indicators between male and female partners, and (3) explore attributes of discordance between self vs. proxy spousal reports by including both partners' characteristics, using multilevel hierarchical models and structural equation models. The paper discusses implications of findings for the survey measurement of alcohol use and for future research.

A Management Model for Continuous Data Collection:

Reflections from the National Survey of Family Growth, 2006-2010

Nicole Kirgis, University of Michigan, Institute for Social Research (nkirgis@umich.edu); Jim Lepkowski, University of Michigan, Institute for Social Research (jimlep@umich.edu)

In its seventh cycle, the National Survey of Family Growth (NSFG) changed its design from a periodic, large, one-time data collection to a continuous data collection over the course of several years. The goal of the continuous design was to interview a large, nationally representative sample of men and women 15-44 years of age with greater cost-efficiency and greater control over sample size, data quality, and cost through the use of a more efficient sample design, and extensive use of paradata to make real-time management of interviewer effort possible.

Continuous data collection in the NSFG context is a design of on-going sampling of the U.S. age-eligible household population so that interviewing and data production occurs daily, with changes in questionnaires and sampling areas annually. This design seeks to maximize the use of resources and obtain high-quality responses by (1) Producing a balanced, continuous work flow to maximize efficiency, (2) Using a small, cross-trained project team, and (3) Monitoring data collection in real-time in order to make adjustments to the design.

This paper summarizes the design and management strategies used for data collection in a continuous environment. Specifically, the paper will highlight the sampling strategies, paradata monitoring techniques, and responsive design initiatives used for the successful operation of the continuous National Survey of Family Growth.

Using the Generic Ballot to Predict Seat Change in Midterm Elections

Jeffrey M. Jones, The Gallup Poll (jeff_jones@gallup.com)

The generic ballot has proven to be an accurate predictor of the midterm election vote in past elections. However, the key result of the elections is the change in party control of seats in Congress. Gallup has developed a regression model that allows the generic ballot result to be used to predict the distribution of seats by party as a result of the elections. For 2010, the model suggests the Democrats could still lose the popular vote by as many as four to five percentage points but maintain majority control of the U.S. House of Representatives in the 2010 elections.

The Theory and Evaluation of Anchoring Effect in Visually Administered Extremely Long Response Lists

Kumar Rao, Gallup Inc. (kumar_rao@gallup.com)

Numerous studies (Payne 1971; Krosnick and Alwin 1987; Ayidiya and McClendon 1990; Schwarz, Hippler, and Noelle-Neumann 1992; Malhotra 2009; Malhotra 2008) have examined response order effects in visually administered questionnaires. A number of theories have also been proposed to explain this effect, with the most popular one being the theory of satisficing (Krosnick and Alwin 1987; Krosnick 1991). According to this theory, when response alternatives are presented visually, respondents begin at the top of the list and consider each alternative individually, all while establishing a cognitive framework for evaluating later response alternatives. Thus, there is a bias toward selecting earlier response alternatives in such lists. In this paper, I argue that, for extremely long response lists (i.e., one that spans an entire page in a mail survey), this heuristic approach as proposed by the satisficing theory may not be applicable and that the sheer nature of the task at hand may in fact dictate a different approach. Under these circumstances, respondents are likely to follow a search heuristic that involves the use of alphabetization and memory in order to select response alternatives, and therefore, no bias towards earlier items in the list will occur. In this study, I propose a theory of response order effects that does not produce primacy effects proposed by the theory of satisficing. I draw data from multiple surveys to test my predictions made by the theory about the expected nature of response order effect. I discuss the findings from the study and conclude with recommendations for future research.

The Effectiveness of Stratification by Cellular Switch Points: The Minnesota Experience

David Dutwin, Social Science Research Solutions (ddutwin@icrsurvey.com); Kathleen Call, University of Minnesota (callx001@umn.edu); Susan Sherr, SSRS (ssherr@ssrs.com)

Stratification of landline telephone exchanges to increase the probability of locating respondents clustered by location, race and ethnicity, income and/or other measures is common in survey practice today. Unfortunately, the wealth of information provided by the major survey sampling providers at the landline telephone prefix level does not extend to cell phone prefixes. However, there is the possibility of approximating landline exchange-level stratification within cell phone samples by stratifying cell phones by cellular switch point. The 2009 Minnesota Health Access Survey disproportionately stratified its landline sample into thirteen geographic strata and oversampled four strata relative to the other nine. The cell phone sample was flagged by switch point and sampled proportionately. Post hoc analyses created strata in the cell phone sample that mirrored the stratification in the landline sample and explored the likelihood that sample from specific switch points fell into expected strata. Results indicate that switch points are an effective means of stratifying sample at the state level, with certain significant limitations for both highly urban and highly rural areas of the state.

Surveying a Unique Population: The 2009 National Survey of African American Churches
Sara Bausch, ICF Macro (sara.h.bausch@macrointernational.com); Matt Denker, ICF Macro (matthew.j.denker@macrointernational.com)

Including members of minority groups in study populations is critical to research hoping to address disparities in health and social welfare. Investigators engaged in this research often encounter distrust among minority populations, particularly African American participants. As mainstays in the community, investigators increasingly look to the African American church as a way to reach participants, and as such, contact and engagement of pastors and other church leaders is critical to success.

In the spring and summer of 2009, ICF Macro had the unique opportunity to survey leaders in African American churches across all denominations. This study presented several challenges. This paper details the issues around sample frame development, response rate, cooperation rates, and other logistical issues encountered during this study. The paper is mainly concerned with lessons learned for additional studies involving minority affiliated churches and community organizations and offers strategies for contacting and engaging church leadership.

A Moving Target: The Effect of Changing Respondents in a Panel Survey of Households
Kate E Bachtell, NORC at the University of Chicago (dalton-kate@norc.org); Beth Fisher, NORC at the University of Chicago (fisher-beth@norc.org); Ned English, NORC at the University of Chicago (english-ned@norc.org); Catherine Haggerty, NORC at the University of Chicago (haggerty-cathy@norc.org)

Making Connections is a study of ten urban communities throughout the United States, funded by the Annie E. Casey Foundation. The study employs a unique hybrid sampling design that combines a cross-sectional focus on neighborhoods and a panel survey of households with children. The major advantage of this design is that the data may be used to measure the effects of initiatives on both neighborhoods and families at several points in time. This paper offers an in-depth analysis of one particular element of the sample design: the decision to reassess household composition in every wave of data collection and potentially change respondents within the same family.

In the baseline, we employed a semi-random approach that prompted interviewers to use a Kish table to select an adult to act as the respondent in adult-only households. In households with children, interviewers randomly selected a minor to serve as the “focus child” and then asked to speak with the parent or guardian who knew the most about the focus child (meaning the respondent selection was non-random). In Waves 2 and 3 the screening procedure was driven by two main criteria: 1) whether or not there were children in the household at the time of the interview (the “current household”) and 2) whether or not the current household composition had changed since the previous wave. Using the data from three waves (gathered in 2002-2004, 2005-2007, and 2008-2009), we examine the extent to which switching respondents produced substantive changes in the survey data. We draw from prior comparisons of respondent selection procedures (Oldendick et al 1988, etc.) to investigate the tradeoffs associated with semi-random respondent selection. Our discussion tackles the difficult question of how much survey data reflect the perceptions of behaviors and conditions rather than the experiences themselves (Duncan and Kalton 1987, 109).

How Valid Are Cross-Cultural Differences in Sexual Behaviors and Sexually Transmitted Diseases? Using Cognitive Interviews to Examine Odd Data Patterns
Alisu Schoua-Glusberg, Research Support Services (alisu@researchsupportservices.com); Heather Ridolfo, National Center for Health Statistics (evd7@cdc.gov)

Cognitive interviewing is often used to evaluate questions and make changes prior to fielding a survey. However, increasingly the method is being used to assess the quality of existing survey data. For example, in previous years of the National Health and Nutrition Examination Survey, unlikely response patterns in the data were found. Hispanic, particularly Spanish-speaking Hispanics, were found to be less likely than other racial/ethnic groups to report having ever engaged in sexual behavior, and more likely to respond don't know or refuse to questions asking about sexual identity. These findings indicated potential translation and/or interpretation problems among this subgroup of the sample. In order to address this, the National Center for Health Statistics conducted a cognitive interviewing project which consisted of 58 interviews with 28 English speaking adults and 30 Spanish speaking adults. In this paper we present findings from this project and demonstrate how this method can be used to understand patterns found in the quantitative data.

Surveys of Older Adults in the US: Technology Matters

Kelly Quinn, University Of Illinois at Chicago (kquinn8@uic.edu)

The dynamic demographic structure of the US emphasizes continual revaluation of its subpopulations and subcultures to surface potential systemic biases in survey research. These examinations often consider differences related to regional language and cultural practices, such as those found in immigrant populations; often overlooked however are considerations related to age and aging. As a sub-segment that is not only increasing as a percentage of the total population, but also one that carries significant political and economic weight, adults over age 55 have markedly different characteristics than the general populace. The accommodations made by many survey researchers for this group are limited to the physiological and psychological factors associated with declines in cognitive functioning and health. While these considerations are significant, they comprise a limited view which fails to recognize distinct cultural practices and values. One important area that lacks scrutiny is the adoption and use of newer information and communication technologies; upon closer examination, adoption practices also surface cultural value distinctions related to mediated social interaction.

Lower participation rates by older adults in newer technologies are often treated as a literacy issue by many survey researchers, and are compensated by an adoption to paper-and-pencil or face-to-face survey formats. But these approaches diminish the difference as a proficiency dichotomy, and overlook emerging cultural values related to mediated communications. When viewed culturally, differences in the adoption and use of technologies by older adults raise important considerations regarding the use of more technologically-dependent modalities and cultural perceptions regarding privacy may impact the appropriateness of their use. The purpose of this paper is to review the literature concerning the more conventionally-understood methodological issues inherent in surveying older adults, and to additionally to offer evidence of some lesser-recognized cultural considerations present in this subpopulation.

Minimizing Fieldwork Costs in National Random Door-to-Door Surveys

Andrew I Zelin, Ipsos MORI (andrew.zelin@ipsos.com)

Door-to-door interviewing is still very much a mainstay of sampling in the UK and developing countries for studies where a high degree of robustness is required. This paper describes a fast and practical method for batching-up randomly-selected sampling points into "clusters" of three; such that the points in each cluster are closely-positioned. The sample remains technically unclustered, and therefore benefits from the statistical robustness and power of pure random samples, yet to fieldwork departments, these can be seen as "clustered" and they can thus benefit from cost savings in terms of reduced costs of travelling within a cluster. The work centers around a large-scale national door-to-door Media survey and the

programming is carried out in SAS. The aim is to position as high a proportion of sampling points within a survey region as possible into “triples” such that no two points within a triple exceed a certain distance from each other. Described are the methods for allocating points to these triples on nearest-neighbor bases, whilst making suitable adjustments to the distance measures in order to avoid being left with the challenging task of trying to group the most peripheral points. Further methods are employed to optimize the chance of positioning the residual singles and pairs into triples. The superiority of this method over a more manual process of tripling the points is also confirmed, in terms of the mean travel distance required of fieldwork staff within a triple. A further advantage of this method is its ability to be carried out based on just the northings and eastings of the original sample point locations, obviating the need for slow and expensive specialist mapping software to be used and purchased. The method is flexible and can easily be bolted on to in-house sampling programs.

AAPOR Concurrent Session B

Thursday, May 13, 2010

4:00 p.m. - 5:30 p.m.

AAPOR Task Force Report on Surveying Cell Phones in the U.S.

2010 AAPOR Task Force Report on Surveying Cell Phones in the U.S.

Paul J Lavrakas, Independent Consultant (pjlavrak@optonline.net); Chuck Shuttles, The Nielsen Company (Chuck.Shuttles@nielsen.com); Linda Piekarski, Survey Sampling (linda_piekarski@SurveySampling.com); Charlotte Steeh, CDC (steehc@bellsouth.net); Scott Keeter, Pew (skeeter@pewresearch.org); John Hall, Mathematica Policy Research (JHall@mathematica-mpr.com); Howard Fienberg, CMOR-MRA (howard.fienberg@mra-net.org); Anna Fleeman-Elhini, Arbitron (anna.fleeman-elhini@arbitron.com); Tom Guterbock, U. of Virginia (tmg1p@virginia.edu)

This session will be a structured roundtable format. The session will present and discuss the 2010 report of the reconstituted AAPOR Cell Phone Task Force. The report, written by the 21 members of the Task Force, is anticipated to be released by AAPOR Council in spring of 2010. Each of the chairs of the seven subcommittees that wrote the various sections of the reports will present a five minute summary of their respective subsections. These sections address cell phone surveying issues of (1) Coverage and Sampling, (2) Nonresponse, (3) Measurement, (4) Weighting, (5) Legal and Ethical Considerations, (6) Operations, and (7) Costs. After the summary presentations, the roundtable session will be opened up to discussion among the presenters and the audience. Another member of the Task Force will moderate the session. The Task Force chair will close the session with a summary of the report and the discussion, as well as identifying existing knowledge gaps.

Changing Attitudes and the Intensity of Public Opinion

Framing and Context Effects in Support for Same-Sex Marriage and Civil Unions

Jennifer Heerwig, NYU (ennif@nyu.edu); Brian J. McCabe, NYU (bjmcc@nyu.edu)

Over the past fifteen years, the legal recognition of same-sex relationships has emerged as one of the most controversial and oft-debated social issues. As the public debate gained traction after the 2003 Massachusetts Supreme Court decision legalizing same-sex marriage, researchers and polling organizations increasingly sought to measure Americans' attitudes on the topic. Unfortunately, these polls are beset by substantial heterogeneity in polling frames and question context. Over the last decade, polls have framed the issue as one of gay marriage, same-sex marriage and homosexual marriage, but little research evaluates the effect of shifting frames on public attitudes in this policy domain. In this research paper, we report the results of an Internet experiment testing for framing and context effects in public attitudes towards the legal recognition of same-sex relationships. We report no difference in mean support when respondents are asked about gay, same-sex or homosexual marriage and civil unions, but observe substantial variation in the strength of opposition based on these frames. Notably, we find this effect concentrated among middle-aged and older respondents. We also test the impact of context on reported attitudes by randomly alternating the order of the marriage and civil unions questions. We report

an increase in support for civil unions when asked after the question about marriage, but do not find statistically significant context effects for marriage. We discuss the implications of our findings in the context of ongoing efforts by advocacy organizations to reframe the debate on same-sex marriage.

Technological Innovation in America's Backyard: The Intersection of Media Use, Political Partisanship, and Public Opinion of Biofuels

Nicholas Brigham Schmuhl, University of Wisconsin-Madison (nbschmuhl@wisc.edu); Michael A. Cacciatore, University of Wisconsin-Madison (mcacciatore@wisc.edu); Andrew R. Binder, University of Wisconsin-Madison (arbinder@wisc.edu); Bret R. Shaw, University of Wisconsin-Madison (brshaw@wisc.edu); Dietram A. Scheufele, University of Wisconsin-Madison (scheufele@wisc.edu)

Research has identified several demographic and cognitive variables that contribute to attitudes toward technological innovations. Recently, studies have associated factual knowledge with negative perceptions of science and technology. Others have found that media play a role in shaping public opinion regarding complex scientific and technological issues; however, the public are not passive receivers of information, and certain perceptual filters may lead audience members to interpret media messages differently and thus reach different conclusions about whether to support innovation. With this paper, we explore these mechanisms in the context of biofuels, an alternative energy innovation that could have profound impacts on the everyday lives of American citizens. Such alternative energy solutions depend upon public support for use and development, and public opinion can make or break this technology in the changing energy market.

To investigate these relationships, we analyze data from a statewide RDD survey, conducted between April and June 2009, of 600 individuals in a top biofuels-producing state. We assess relationships between (1) support for biofuels in terms of societal and economic impacts and (2) knowledge, media use, and partisanship in a social context defined by rural American values as well as strong traditions of agricultural research and higher education. Our results suggest that political partisanship and knowledge directly affect overall support for biofuels. Meanwhile, attention to certain types of news media affects our two dependent variables (support based on societal or economic impacts) differentially. However, we find that partisanship moderates the effects of attention to particular media sources and content areas – especially political television – on support for biofuels, with support increasing among Democrats but decreasing among Republicans as their attention to political television increases. We discuss the implications of these interactive relationships as well as other demographic and cognitive characteristics and their effects on support for biofuels.

Measuring Intensity of Opinion

David W. Moore, University of New Hampshire (dmoore62@comcast.net); Andrew E. Smith, University of New Hampshire (andrew.smith@unh.edu)

Over sixty years ago, the “founding fathers” of public opinion research recognized the importance of measuring intensity of public opinion. As Daniel Katz wrote in 1944 in *Gauging Public Opinion*, “to interpret poll results adequately it is necessary to know whether an expressed attitude represents a superficially held view which may be discarded the next moment or whether it reflects a cherished conviction which will change only under unusual pressure.” He wanted to know “whether or not an individual with a given opinion holds that opinion strongly enough to take the trouble to go out and vote for it or fight for it.”

Since then, there has been little research on how to measure such intensity, especially to determine whether the intensity is sufficient for a person to either vote or fight for that view. Many pollsters do occasionally ask whether people feel “strongly” or “not strongly” (or “moderately”), and some also occasionally ask if they would vote for (or against) their representative on the basis of that issue. But for different reasons, these measures have not proved to be very helpful.

At the 2002 AAPOR annual conference, in their paper “Directive vs. Permissive Public Opinion” (updated in 2007), Moore and Jeff Jones reported on another method: asking if respondents would be “upset” if the opposite view that they had just expressed were to prevail.

The proposed paper for the 2010 conference would report on additional experiments conducted by the UNH Survey Center, which examine how the three measures correspond with each other – whether people who feel “strongly” are the same as people who would be “upset” and who would “vote” for/against a candidate on the basis of that issue. The paper also suggests which measure provides the most meaningful result for political leaders and the public.

Parental Consent and Changing Abortion Attitudes, Examining Trends in California

Mark Baldassare, Public Policy Institute of California (baldassare@ppic.org); Jennifer Paluch, Public Policy Institute of California (paluch@ppic.org); Sonja Petek, Public Policy Institute of California (petek@ppic.org)

Public opinion about abortion policy has played a decisive role in national and state politics for decades. Since abortion was legalized in the U.S. through the landmark 1973 Roe v. Wade court decision, numerous states have enacted restrictions limiting abortion. Currently, a majority of states require parental involvement, taking the form of either notification or consent in a minor child's decision to have an abortion. In November 2008 California voters defeated a parental notification measure (52% to 48%) by a relatively narrow margin with a deeply divided electorate for the third time in three years (53% to 47% in 2005; 54% to 46% in 2006). This will likely not be the last time California voters weigh in on parental notification, as signature gathering is underway for another parental notification initiative in 2010. Despite California's reputation as a solidly pro-choice state and rejecting parental notification measures, the Public Policy Institute of California (PPIC) Statewide Survey this year found majority support for the idea of parental notification and an increase in favoring greater restrictions on the availability of abortion. The paper will examine whose attitudes are shifting and whether this shift could result in changing public policy in a state where voters make policy decisions at the ballot box.

Using data collected as part of the PPIC Statewide Survey, this paper analyzes trends in public opinion from 1998 to 2009 on abortion and parental notification issues, including attitudes towards Roe v. Wade, government restrictions and access to abortion, personal beliefs and moral objections to abortion, preferences on reducing abortions, individuals' reasons for voting for and against 2008's parental notification measure, and perceived importance of the vote outcome. It will compare trends in California over time with national surveys, and provide a multivariate analysis to determine political and demographic factors associated with abortion attitudes.

A Network Analytic Approach to Understanding Cross-Platform Audience Behavior

Thomas Ksiazek, Northwestern University (tom.ksiazek@gmail.com)

This methodological paper explains and implements a network analytic approach to the study of cross-platform audience behavior. It begins by conceptualizing large-scale patterns of media consumption in network terms, treating media outlets as nodes and the levels of audience duplication among them as relations, or ties. Following that, it explains

two common measures of audience duplication, Both Duplication and Primary Duplication, and offers a new measure, Deviation-from-Random Duplication. In doing so, techniques for converting duplication data into network data are discussed. Potential research agendas are offered to illustrate the value in adapting network analysis to the study of audience behavior. Using data from Nielsen's TV/Internet Convergence Panel, these studies analyze patterns of audience fragmentation and polarization, and identify "media publics," or audience segments defined by common consumption patterns across platforms.

Cross-national Attitudes & Beliefs

Fifty Years After "The Civic Culture": Insights From the Gallup World Poll 2008-2009

Rene Bautista, University of Nebraska-Lincoln (rene.bautista@huskers.unl.edu); Ana Lucia Cordova Cazar, University of Nebraska-Lincoln (analucordova@yahoo.com); Allan McCutcheon, University of Nebraska-Lincoln (amccutch@unlserve.unl.edu); Tarek Albaghal, University of Nebraska-Lincoln (mtalbaghal@gmail.com); Lei Huang, University of Nebraska-Lincoln (leihuang03@hotmail.com)

Almond and Verba's groundbreaking cross national survey, "The Civic Culture" (1963) focused on the long-debated concept of political culture in five countries. In "The Civic Culture Revisited" (1980), several scholars argued that political culture is a concept subject to change over time across different countries mainly due to changes in social, political and economic conditions. Fifty years after Almond and Verba's seminal study, the present paper offers a comparison of the countries that served as basis for the original study using data from the most recent Gallup World Poll (2008 and 2009). The Gallup World Poll is a probability survey that uses a standard set of core questions across countries. On average, 1,000 cases are collected in each country. Seven Gallup World Poll indices, namely Citizen Engagement, Law and Order, Diversity, Personal Economy, Religiosity, Corruption, and National Institutions, are examined to shed light on theoretical concepts pertaining to political culture. These indices assess a number of values that have been theorized as important in democratic stability: tolerance, freedom of speech, trust in government, and participation in voluntary associations, among others. Regression analyses are conducted to examine correlates of Citizen Engagement. Preliminary analyses suggest that the United States, United Kingdom, and Germany have the higher levels of political engagement, net of other factors. Italy and Mexico, on the other hand, continue to have lower levels of political engagements. Respondent's education appears to play the strongest role predicting citizen engagement in all five nations.

-Almond, Gabriel and S. Verba (1963), *The Civic Culture*, Princeton University Press

-Almond, Gabriel and S. Verba, eds., (1980), *The Civic Culture Revisited*, Sage Publications

Attitudes Toward Democracy Twenty Years After the Fall of the Berlin Wall

Juliana Menasce Horowitz, Pew Research Center (jhorowitz@pewresearch.org); Richard Wike, Pew Research Center (rwike@pewresearch.org)

Twenty years after the fall of the Berlin Wall, publics in the former Eastern bloc express less enthusiasm for democracy than they did as the Soviet Union was collapsing. This paper will explore attitudes toward democracy in eight Eastern European countries, using data from two unique public opinion surveys: a 1991 survey by the Times Mirror Center (the forerunner of the Pew Research Center for the People & the Press) and a fall 2009 survey by the Pew Research Center's Global Attitudes Project. The countries included in the study are Bulgaria, Czech Republic, Hungary, Lithuania, Poland, Slovakia, Russia, and Ukraine.

This paper will use a multivariate model to test a number of hypotheses related to attitudes toward democracy in the eight former communist countries surveyed. First, we expect views of democracy to be more positive among those who have embraced free market capitalism. Second, we theorize that those who are more engaged in civil society have more favorable attitudes toward democracy. Third, we expect economic prosperity to have a positive effect on view of democracy. And finally, we hypothesize that negative evaluations of how democracy is working in each of the countries surveyed reflect a gap between what people want from democracy and what they currently have. We will also explore the relationship between support for democracy and demographic variables such as education, age and religiosity.

Cross-Cultural/National Issues from Implementing the Global Adult Tobacco Survey in 14 Countries

Jeremy Morton, Centers for Disease Control and Prevention (jmorton@cdc.gov); Krishna Palipudi, Centers for Disease Control and Prevention (kpalipudi@cdc.gov); Samira Asma, Centers for Disease Control and Prevention (sasma@cdc.gov)

The purpose of this paper is to describe the various cultural/national issues and challenges that were experienced in implementing the Global Adult Tobacco Survey (GATS) in 14 countries. GATS was developed to provide a global standard protocol for consistent monitoring of adult tobacco use and was designed to produce national and sub-national estimates across countries and indirectly measure the impact of tobacco control and prevention initiatives. GATS is a nationally representative household survey of all non-institutionalized, men and women age 15 years and older using a standard core questionnaire, sample design and data collection methods (CAPI using handheld computers) that were reviewed and approved by experts both from developed and developing countries across the world in tobacco control, survey design and methodology.

Initially GATS was conducted in 14 countries in 2008-2009: Bangladesh, Brazil, China, Egypt, India, Mexico, Philippines, Russian Federation, Thailand, Turkey, Ukraine, Uruguay, and Vietnam. Implementing GATS in these different countries presented cultural/national challenges related to sample design, questionnaire adaptations, language translation, training, electronic data collection, fieldwork implementation, and data release. This paper will discuss these issues and lessons learned that will be important for GATS in the future and perhaps might benefit other global surveys.

Public Opinion and Crisis: Insecurity and Economic Situation in the Perception of the Mexican Public Opinion 2007-2009.

Paul Francisco Valdes Cervantes, Parámetro Investigación (paul.valdes@gmail.com); Nelson Arteaga, UAEM (arbnelson@yahoo.com)

The importance of the major themes that dominate the public agenda in Mexico is public insecurity and economic situation, and their influence on the performance of authorities and the level of party identification of the three main political parties.

Main social concerns are reflected in the public agenda that does not necessarily coincide with the media agenda and the political agenda. During recent years, issues related to public security and economic situation were the themes that have dominated the national agenda.

It reviewed the approaches of the public agenda and public opinion. For most authors, the formation of public opinion is flowing from top to bottom, elite to the mass society, where the media serve as a transmission belt for framing messages (framing and priming). On the other hand, less frequently may be the formation of the view from the bottom up (bubble up).

Is seeking evidence on the possible influence of the public agenda, particularly in the evolution of the performance of authorities and the level of party identification in the period October 2007 to March 2010.

Effect of Religion on Other Social & Political Attitudes

Diversity and a Sensitive Social Issue: Hispanic Catholics and Their Comparative Response to the Clerical Abuse Crisis

William C McCreedy, Knowledge Networks (bmccready@knowledge networks.com); Mark Gray, Center for Applied Research on the Apostolate (CARA) (mmg34@georgetown.edu); Jeffrey Shand-Lubbers, Knowledge Networks (jlubbers@knowledge networks.com)

This paper is based on a survey commissioned by the US Catholic Bishops to examine the attitudes of Catholics toward the clerical sex abuse crisis. Two methodological issues are addressed: the accurate representation of the opinions of the growing number of Hispanic Catholics and the control of potential social desirability bias present in an interviewer-conducted survey about a sensitive topic. The combination of these two needs dictated using a self-administered survey with a probability-designed Internet sample that included the proper proportion of US Hispanics across the language spectrum, from those that speak predominantly English to those that speak predominantly Spanish and all variations between.

A key element in the study is the accurate reflection of the responses of the Hispanic/Latino Catholic population in the US since this group represents an increasing proportion of the US Catholic population and it is inappropriate to survey the national Catholic population without taking cognizance of this rapidly changing demographic. In addition, since the topic of this study is public reaction to how the Church handled the clerical sexual abuse crisis, there was a need to have a regionally diverse sample since the crisis was of different magnitudes in different regions.

The sample is just under 2,000 cases with subsamples in three regions and a high-quality sampling of Latino Catholics with the survey administered in both Spanish and English. The analysis focuses on comparisons of attitudes toward the abuse crisis between English-dominant Hispanic Catholics, Spanish-dominant Hispanic Catholics and the non-Hispanic Catholic majority. There are also inter-regional analyses demonstrating how the response differed in varying Church locales. The sexual abuse crisis has cost the Catholic Church dearly in terms of both financial resources and moral credibility. This survey provides data about the diverse public reaction to the crisis and the Church's response to it.

Differences Between Self-Identified "Born-Again" and "Evangelical" Christians in Sub-Saharan Africa

Scott F Clement, Pew Research Center (sclement@pewforum.org)

Given the substantial influence of religion in politics in several countries, accurately measuring religious groups in public opinion surveys is of paramount importance. In particular, differences among "born-again" and "evangelical" Christians have led to a lively debate among pollsters and scholars alike. While much has been written on this subject in the United States, the issue remains under-explored internationally. Using survey data from 17 sub-Saharan African countries that include questions regarding religious beliefs as well as social and political attitudes, we analyze the differences in the religious, political and social profile of those who identify their Christian identity as "born-again" versus those who describe themselves as "evangelical." We find that in eleven countries a significantly higher proportion of

respondents identify themselves as “born again” rather than “evangelical” Christians, further highlighting the need to understand the differences among these two groups.

Further, we explore in detail born-again and evangelical identification among Catholics, analyzing how this phenomenon compares with previous research on Catholic evangelicalism in the United States. In particular, we find that in several sub-Saharan African countries, a majority of Catholics identify as born-again Christians, and that substantial minorities identify as evangelicals. By comparison, previous research has found that only one-in-six American Catholics identify as either born-again or evangelical.

Who’s Lying? A Cross-National Comparison of the Correlates of Over Reported Religious Service Attendance

Philip Brenner, University of Michigan (pbrenner@umich.edu)

That Americans overreport their religious service attendance is well accepted. Comparisons of rates of religious service attendance garnered from stylized questions with those from time diaries strongly support this understanding. However, where previous work focused only on recent American data, a wider net is cast in this paper, including data from Canada and twelve European countries, and from the past five decades. Moreover, this line of research is further extended beyond comparisons of the marginal rate of overreporting by-country (Brenner 2009) to investigate the locus of overreporting. A series of logistic regression models are estimated, predicting both measures of attendance using a suite of key independent variables widely acknowledged in the existing literature to be highly associated with attendance: gender, age, marital status, presence of children in the household, income, and education. The effects of these covariates are compared between models for each of the dependent variables to answer the following question: can these factors explain overreporting where and when it occurs? However, direct comparison of logits between models may be problematic due to the often unwarranted assumption of equal residual variation between logistic regression models. In testing these assumptions and making comparisons, the work of Allison (1999), Hoetker (2004, 2007) and Williams (2008) is highlighted. Findings suggest that none of the demographic predictors of self-reported attendance explains where overreporting occurs at high rates (i.e., the United States) and where it occurs at low rates or does not occur at all (i.e., countries in Northern and Western Europe). Rather, findings suggest that overreporting is better attributed to the directness of the stylized measure (commonly used in conventional survey questionnaires) and the resultant identity-priming effect of the directive stylized question.

“The Times They Are A-Changin’”? Ethnic and Religious Diversity and Their Effects on the Social, Political and Economic Beliefs and Attitudes of Young People in England

Nick Moon, GfK NOP Social Research (nick.moon@gfk.com); Iain Noble, University of Westminster (I.Noble@westminster.ac.uk)

The UK is one of the most ethnically diverse states in Europe.

Members of ethnic minorities are now a significant proportion of the population, and importantly, the proportion of ethnic minorities is higher among younger age groups. Diversity is here, has increased and will continue to increase for the foreseeable future.

One effect of this increase in ethnic diversity is an increase in religious diversity, in particular the creation of a significant minority of Muslims. This is partly the result of demographic change but also because the children of Muslims are much more likely than the children of Christians to profess the parental religion.

These changes have led to increasing concern in the UK about ‘community cohesion’, and the ‘living of separate lives’ by some minorities. Specific concern has been raised about the possible disaffection of certain groups, especially younger black males and young Muslims.

Traditionally in the UK differences in social and political attitudes are associated with age, gender, education and, especially, social class. Most national surveys of such attitudes in the UK do not have ethnic minority samples that would permit robust analysis of differences within younger age groups that might be associated with ethnic or religious factors. There are however reasons to do so. The experience of younger members of the minorities is likely to have been substantially different from their parents and even from that of their older siblings.

We use data from a large scale longitudinal study of young people, gathered through interviews between the ages of 13 and 18, to explore these issues. We conclude that there is little sign of general effects but that there are specific signs of estrangement in black males from economically disadvantaged backgrounds and some evidence of divergence among Muslims on issues linked to personal behaviour and responsibility.

Conflict, Integration or Independence: Public Attitudes on Science and Religion in American Society

Rob Suls, Pew Research Center (rsuls@pewresearch.org); Daniel Cox, Public Religion Research (dcox@publicreligion.org)

Science and religion each offer distinct ways of understanding the world. Each purports to offer truth about questions important to humankind—such as where did we come from—and in some cases they take us in entirely different directions. Yet how often are science and religion at odds with each other, and how do we view this conflict?

Perhaps in no society is the potential for conflict between religion and science more apparent than in the United States. It has frequently been observed that Americans maintain high levels of traditional religious commitment particularly when compared with other industrialized nations. More than eight-in-ten Americans are affiliated with a particular religious tradition and a solid majority (56%) say religion is ‘very important’ in their life.

Yet despite the importance of religion for many Americans, scientific progress and technological development have been hallmarks of American society. And public views of science and scientists are overwhelmingly positive. Seven-in-ten Americans say that scientists contribute a lot to the well being of society and 84% say that science has had mostly a positive effect on our society.

In this paper we will examine how religious affiliation and beliefs influence views about science and how the opinions of these diverse communities inform a wide range of science policy alternatives. We will pay particular attention to issues where science and religion have thought to be at odds, for instance in the case of evolution. This paper will draw on two new datasets from the Pew Research Center and American Academy for the Advancement of Sciences, which includes a survey of the general public and American scientists. Each dataset includes a wealth of measures on a wide range of science topics, attitudes about science’s role in society as well as measures about conflict between the two.

Identifying & Correcting Non-response Bias

Using Post-Stratification Weighting to Reduce Nonresponse Error in Student Survey Data: Preliminary Evidence

Matthew W. Courser, Pacific Institute for Research and Evaluation (mcourser@pire.org); Stephen Shamblen, Pacific Institute for Research and Evaluation (sshamblen@pire.org); Joel Grube, Pacific Institute for Research and Evaluation (grube@prev.org)

Surveys of student populations provide important data to policymakers and researchers on underage alcohol and drug usage, risky student behavior, and school safety. Because student respondents usually are minors, informed consent from parents is a required component of these surveys. A large body of research has found that the parental consent process can result in student survey data having large amounts of nonresponse error—especially when active parental consent procedures are used (Courser, et al., 2009).

Researchers have worked to develop strategies to reduce nonresponse error in student survey data. To date, student, parent, and teacher incentives and multiple follow-ups have been shown to reduce nonresponse error in student surveys. However, these strategies require substantial financial resources and extra time, which many student survey efforts (particularly those funded by schools themselves or by local communities) do not have.

Post-stratification weighting techniques have been used successfully in many types of surveys and in some public health studies to adjust survey data to better reflect the underlying population parameters. However, to date no published studies report using post-stratification weighting to reduce nonresponse error in student survey data.

Our paper uses data from a student survey conducted during the 2007-2008 school year with 14 Kentucky school districts and investigates whether post-stratification weighting can be used as a low-cost technique to reduce non-response error. Because 7 of our school districts implemented the survey with passive consent procedures and 7 implemented with active consent procedures, we will examine whether the utility of weighting techniques vary across consent conditions. We discuss our findings and their practical implications, with particular emphasis on whether the student population data maintained in standardized school databases is sufficient for the effective use of post-stratification weighting.

Nonresponse Bias Correction in Telephone Surveys Using Census Geocoding: An Evaluation of Error Properties

Paul P. Biemer, RTI International (ppb@rti.org); Andy Peytchev, RTI International (andrey@umich.edu)

The threat of nonresponse bias has been increasing with the precipitous decline of response rates in random-digit-dialed (RDD) telephone surveys. Often, researchers have only geographic information for sample cases. Census demographic information can be appended at varying levels of geographic aggregation for both respondents and nonrespondents in landline samples, in an attempt to correct for nonresponse bias. The effectiveness of this method depends on the error properties of the census geocoding (CG) process, yet it has not been thoroughly evaluated. At an extreme, error in the CG process can do more harm than good for survey estimates. More realistically, components of the process can be identified as more susceptible to error and improvements can then be made.

The CG process can be implemented in various ways, such as identifying census block groups for listed telephone numbers and census tracts for unlisted numbers. Matches to geographic areas may be erroneous because of incorrect addresses, and also from different postal and census geographic definitions. Although rates of some types of mismatches have been examined, their impact on estimates

has not. Furthermore, demographic information in smaller geographic areas may not match the demographic sample composition, inducing higher variance in adjusted estimates without necessarily reducing bias even in demographic variables.

Using parameters from an RDD survey, we impose nonresponse on a face-to-face study that allows us to decompose the error in the CG process. This approach provides a gold standard for respondents and nonrespondents, as well as for listed and unlisted telephone numbers. Preliminary findings show that while for some variables incorrect matches of unlisted numbers contribute the most to bias in the CG process, surprisingly, bias is also relatively large for correctly matched telephone numbers. We continue this research and aim to offer practical suggestions for the use of the CG method.

Use of Paradata for Nonresponse Weighting Adjustments

Donsig Jang, Mathematica Policy Research (djang@mathematica-mpr.com); Amang Sukasih, Mathematica Policy Research (asukasih@mathematica-mpr.com); Kelly H Kang, National Science Foundation (kkang@nsf.gov); Stephen H. Cohen, National Science Foundation (scohen@nsf.gov)

In recent years, there has been considerable discussion concerning the difficulty of obtaining high response rates in surveys. The discussion focuses on two general areas—one area focuses on the survey operations point of view where strategies to improve response rates are identified and proposed; the second area assumes operational strategies to improve response rates have been implemented but nonresponse still remains.

In this study, we focus on the second area, which is accounted for through weighting adjustments using data from the sampling frame. In recent years, there has been a growing body of literature that discusses a complex response process to maintain target response rates. The likelihood of responding to the survey depends not only on demographic and socioeconomic characteristics available from the sampling frame but also on other information obtained by data collectors during the field period. This information—known as paradata—includes empirical measurements made during the process of data collection such as interviewer level of effort, type of incentive treatment, etc. Technology has helped data collectors keep track of the information related to the data collection protocol. Statisticians interested in reducing nonresponse bias through weighting adjustments have begun to look at the paradata collected during the survey fielding period and its correlation to response propensity, survey outcomes, or both. Despite growing awareness of the utility of paradata, only a few surveys use these data for weighting adjustments.

Using the 2008 NSRCG data, we will compare response propensity values predicted only with frame data to those with both frame and paradata. We will also estimate key survey statistics using weights based on (1) frame data only and (2) frame and paradata. This empirical study can help us understand the extent of survey bias reduction when combining information about the complex data collection process with frame data for use in nonresponse weighting adjustments.

Who is Missing and How Much Does it Impact Educational Performance Estimates? – Propensity Models Applied to the NAEP Survey Data Merged with the Administrative Data of the High School Transcripts Study

Young Chun, NORC at the University of Chicago (chun-young@norc.uchicago.edu)

This paper examines root causes of nonresponse of 12th graders in the National Assessment of Educational Progress (NAEP), using a model of nonresponse developed by Groves and Couper (1998). NAEP is the only nationally representative and continuing survey of American student knowledge and capability in various subject areas including mathematics and science. NAEP faces the possibility that its

results could be contaminated by a low response rate. In the past 10 years, the overall participation rates at the 12th grade have been 10% to 35% lower than rates at grades 4 and 8. A fundamental issue is how strongly correlated the NAEP variables of interest are with response propensity, the likelihood of participation in NAEP. If socially isolated and poor performing students are less likely to participate in NAEP, student achievement in NAEP is likely to be overestimated. The paper evaluates the statistical impact of nonresponse bias on estimates of educational performance in NAEP, by applying response propensity models to the NAEP mathematics and science survey data and the corresponding school administrative data from over 20,000 seniors in the 2000 High School Transcript Study (HSTS). When NAEP and administrative HSTS data are merged, one has measures of individual- and school-level characteristics for nonparticipants as well as participants in NAEP. Response propensity models are developed by including a social isolation theory-guided set of predictors of NAEP nonresponse. Results indicate that nonresponse was not a serious contaminant, and applying response propensity based weights led to statistically insignificant difference in key education survey outcomes. The results support other recent research (e.g., Curtin, Press and Singer, 2000; Groves, 2006) showing minimal effects on nonresponse bias of lowered response rates.

Examining Nonresponse Bias Using Sample Frame Data

Sunghee Lee, UCLA Center for Health Policy Research (slee9@ucla.edu); Chandra Higgins, Los Angeles County Public Health Department (chiggins@ph.lacounty.gov); Shin Margaret Chao, Los Angeles County Public Health Department (Schao@ph.lacounty.gov); Yvonne Lau, Los Angeles County Public Health Department (yvlau@ph.lacounty.gov)

One way of examining nonresponse bias is to use auxiliary information available in sample frames. The richness of the frame data influences the breadth and depth of the nonresponse bias study. A maternal health survey project, the Los Angeles Mommy and Baby (LAMB) project, was conducted with its sample drawn from birth records obtained from the LA County Automated Vital Statistics System. The frame contains detailed information about the mother, the baby, the birth outcome of the baby as well as some pregnancy-related characteristics of the mother. Questionnaires in English, Spanish and Chinese were mailed to the sampled mothers, and mail nonrespondents were followed-up by telephone interviews. Monetary incentives were offered to participants for completing the survey.

We examine potential nonresponse bias in the LAMB survey with various approaches: 1) by comparing the total sample and the respondents with respect to variables included in the frame; 2) by comparing mail respondents and telephone follow-up respondents with respect to variables asked in the survey as well as frame variables; and 3) by predicting response propensities using frame variables and correlating the predicted propensities with frame and survey variables. This study is expected to provide substantive understandings about exploring nonresponse bias with rich frame data.

Improving Questionnaire Design

Issues in Changing the Measurement Methodology for the Quarterly Tax Survey

Marcus Graham, U.S. Census Bureau (marcus.graham@census.gov); Rebecca L. Morrison, U.S. Census Bureau (rebecca.l.morrison@census.gov); Amy Couzens, U.S. Census Bureau (amy.couzens@census.gov); La Toya Thomas, U.S. Census Bureau (la.toya.thomas@census.gov)

The U.S. Census Bureau has collected tax information from local governments on a quarterly basis since 1962. Based on the Committee on National Statistics' review of the U.S. Census Bureau's state and local government statistics program, the Quarterly Survey of Non-Property Taxes' instrument and sample

design are changing. This paper will provide background about the measurement of local government non-property taxes over time, and then examine the effects of (proposed) changes. These changes include a discussion of two main themes that data collection from governments has in common with other establishment surveys:

- issues with multiple respondents, record retrieval, and terminology that were revealed during cognitive testing and
- collecting data from one target population (tax imposers) versus another (tax collectors)

Both have implications for sample size, instrument design, and data collection costs. In addition, the paper will highlight the importance of identifying the title of the “best respondent” when titles and duties vary across cities, towns, counties, and special districts (e.g., water, schools).

Bento Box Questionnaire Testing: Multi-Method Questionnaire Testing for the 2012 Census of Agriculture

Jaki S. McCarthy, USDA Nat'l Agricultural Stat. Service (jaki_mccarthy@nass.usda.gov); Donald Buysse, USDA Nat'l Agricultural Stat. Service (donald_buysse@nass.usda.gov)

A traditional Japanese bento box is a meal constructed according to principles of balance – separate compartments contain different colors, flavors, and cooking methods to provide an optimum eating experience. As in the bento box, questionnaire testing methods can be combined to provide a balanced questionnaire evaluation. This way, complementary methods can be used to compensate for any methods' shortcomings. For example, analysis of amount of editing in previous data collections may identify questions collecting inaccurate data. These analyses can provide empirical estimates of the size and prevalence of errors, but do not provide information about the reasons for reporting errors. Expert reviews of necessary editing and item imputation or of the questionnaire may elicit potential reasons for respondent errors, but such reviews are quite subjective and inconsistent from expert to expert. In contrast, cognitive interviews can directly examine respondents' response processes and identify areas where respondents may not interpret questions as intended or may not have the requested information available. However, cognitive interviews are not practical for questionnaires with hundreds of variables, and the small, often unrepresentative samples sizes do not provide information on the magnitude of identified problems or how they may ultimately impact survey estimates. In order to overcome the shortcomings of various methods of testing and capitalize on each method's strengths, NASS is conducting a multi-phase, multi-method approach to the revision and testing of the 2012 Census of Agriculture report form. Questionnaire testing will combine information from expert reviews, analysis of data from the 2007 Census of Agriculture, pre-test cognitive interviews, multiple large scale field tests and follow-up cognitive interviews. Each compartment of our testing bento box will be discussed as well as how these methods are balanced to develop the best possible questionnaire for the 2012 Census of Agriculture.

Characteristics of Survey Questions: A Review

Nora Cate Schaeffer, Dept of Sociology, UW Madison (schaeffe@ssc.wisc.edu); Jennifer Dykema, University of Wisconsin Survey Center (dykema@ssc.wisc.edu)

Questionnaire designers are increasingly focusing on the characteristics of questions to better understand how they affect data quality – both measurement error and item nonresponse -- in surveys (e.g., Forsyth et al. 2004; Saris and Gallhofer 2007; Willis 2005). While a number of different characteristics have been studied (e.g. “question difficulty,” “response format/categories”), we are still at the preliminary stages of specifying a comprehensive analysis that catalogues the characteristics of questions and their effects on

both respondents' and interviewers' processing of questions and subsequent data quality. The absence of a comprehensive analysis of the characteristics of survey questions impedes progress on at least the following fronts: Identifying and filling in the gaps in the experimental literature about the impact of question characteristics, focusing cognitive and expert review about question characteristics, research about the effectiveness of such reviews, and training practitioners about how to write survey questions that are likely to yield reliable and valid data. In this paper we compile and review previous analyses of question characteristics. Most of these analyses are implicit, in that the authors produced the analyses in pursuit of some other goal; for example, systems for cognitive review of survey questions are based on an implicit model of which characteristics of questions are likely to affect cognitive processing and subsequently data quality. Some analyses are also ad hoc, in that the authors conducted secondary analyses in which they largely relied on characteristics identified by the researchers who originally collected the data. Based on our description of the features of previous analyses of the characteristics of survey questions, we develop a more comprehensive and systematic description of the features of survey questions that affect the quality of survey data.

Early Stage Scoping: Building High Quality Survey Instruments Without High Costs

Kristin Stettler, U.S. Census Bureau (kristin.j.stettler@census.gov); Fran Featherston, National Science Foundation (f Feather@nsf.gov)

Traditional testing of survey instruments involves drafting an instrument and pretesting it with respondents who fill it out, if possible. New instruments often misfire, as the researcher learns during initial testing that the instrument does not use appropriate terms for the respondent or suffers from a naive view of the subject matter. We highlight our work on a national web survey on R&D activity of 423 state agencies. In this paper, we describe the early stage scoping interviews we used with a "sliding scale approach." This technique allowed efficient and cheap questionnaire development using a total of six one-day visits to five states, supplemented with phone calls to additional states and to earlier respondents.

To assist survey designers in understanding the sliding scale approach, we outline three principles of early stage interviewing. The first principle is "Start fresh." This principle is implemented with a very small number of exploratory interviews to understand the substantive issues from the respondent's point of view. The second principle is to "Learn the respondent's language." During a second small group of interviews, a paper prototype is used to check the language in a draft of the instrument. The third principle is to "Know thy user, for it is not you." The final set of interviews covers usability and navigation issues for a nearly-complete instrument. The sliding scale refers to the changes across pretest interviews -- the portion of the interview devoted to substantive issues starts near 100 percent and diminishes to near zero across the timeline of interviews, while the portion of the interview dedicated to administrative and navigation issues starts near zero and rises to nearly 100 percent. The paper shows how early stage scoping interviews can utilize these three principles to minimize the resources needed to maximize measurement quality while minimizing respondent burden.

Factors in Skip Errors and Omissions on a Self-Administered Paper Questionnaire

Brett McBride, Westat (brettmcbride@westat.com); David Cantor, Westat (davidcantor@westat.com)

Prior research has examined how branching instructions on a self-administered paper survey are associated with rates of skip errors (Redline et al, 2003; Messmer and Seymour, 1982). These skip errors involve respondents both erroneously not answering items and answering items they were not intended to answer. Other features specific to how the questionnaire is designed -- such as response formats, question content, question length and layout -- may contribute to respondents intentionally omitting

answers to questions as well as making skip errors. These Skip Errors and Omissions (SEO) may result from respondent navigational errors, preferences for cognitive short-cuts, confidentiality concerns, or not knowing the answer. Questionnaire design can benefit from understanding the extent to which common design features are associated with SEO.

Data from the 2007 Health Information National Trends Survey (HINTS) was used to test research hypotheses about SEO on a self-administered paper questionnaire. HINTS, sponsored by the National Cancer Institute, included administration of a mail questionnaire sent to a nationally-representative sample of households. The HINTS questions were designed to measure Americans' access to information about health as well as cancer awareness and attitudes.

This paper examines the hypothesis that certain item characteristics in HINTS – item location, response format, and item length – influenced SEO. Average omission rates are calculated for each set of item characteristics and a test of significant differences is carried out against the remaining questions. Rates at which respondents answered items they were not intended to answer are examined. Additionally, this research investigates whether certain respondent characteristics are associated with varying levels of SEO. Regressions using available demographic data as predictors examine the impact of respondent age, sex, and education on aggregate levels of SEO.

Methodological Briefs: Cell Phones, Internet Surveys, and Response Rates.

Completing Web Surveys on Smart Phones

James Dayton, Macro International Inc. (james.dayton@macrointernational.com); Naomi Freedner, ICF Macro (naomi.freedner@macrointernational.com); Kristie Hannah, ICF Macro (kristie.hannah@macrointernational.com)

One-of-eight mobile phones sold this year were smartphones (Internet-enabled cell phones), with sales of these devices growing at a rate of over 20 percent per year. While current smart phone users are demographically quite different from cell phone users in general, these differences will diminish as smart phone penetration exceeds 50 percent over the next five years. Survey researchers are already testing multiple-mode interviewing with cell phones to complete interviews by voice and text message. What about using the internet-enable smart phone to open a new possibility for survey research—text message interviewing? Text message interviewing presents many challenges, from very limited survey complexity to respondent confusion regarding the response the researcher is looking for. A potential solution involves web-based data collection on smartphones using the same web survey tools researchers already have in their data collection toolbox.

The purpose of this pilot is to test if a standard web survey program can effectively be used to administer a web survey through both traditional desk/laptop computers as well as smart phones. IFC Macro is piloting a short web-based survey to cell phone users invited through links provided via e-mail and/or text messaging. Cognitive follow-up interviews will be completed with respondents immediately following the completion of the survey to determine satisfaction and usability levels.

The results of this pilot study will inform survey researchers on the potential barriers to smart phone data collection, including network speeds, screen and text size, navigation challenges, and questionnaire length. Respondent satisfaction with the survey experience will be measured along with willingness to participate in future smart phone survey requests.

By better understanding potential respondents' experience with smartphone survey research, survey researchers can begin to tailor smartphone surveys to maximize the survey experience, thus improving response rates.

Cell Phone Pre-Identification and Geographic Accuracy: Experiences Among Military Personnel
Mark Morgan, Abt SRBI (m.morgan@srbi.com); Daniel Loew, Abt SRBI (d.loew@srbi.com)

Conducting research on cellular telephones has many challenges. Some key concerns are ensuring the safety of respondents, calling at appropriate times of day to minimize disturbing respondents and to maximize the chances of speaking with them and complying with FCC regulations about proper dialing of cellular phones.

This methodological brief offers a pragmatic examination of:

- the accuracy rate of pre-identification tools in classifying cellular phone numbers
- the accuracy level of area code designators in predicting the time zone which governs what time we place calls
- the impact of these factors on call outcomes

The data to be analyzed come from two longitudinal research projects being conducted among members of the U.S. National Guard and Reserves.

Within Household Cell Phone Sharing and other Usage Patterns
Andrew Stavisky, US Government Accountability Office (staviskys@gao.gov); Carl Ramirez, US Government Accountability Office (ramirezcc@gao.gov); James D Ashley, US Government Accountability Office (ashleyj@gao.gov)

Blumberg and Luke, in a methodological brief presented at AAPOR 2009, examined the likelihood of wireless telephone sharing by members within a household. While no direct measurement of sharing behavior was made, those authors found that 63% of households in the National Health Interview Survey reported the same number of people as cell phones, 17% had more cell phones than people, and 20% had fewer cell phones than people. The latter figure suggests a condition that may result in sharing, especially in cell-phone only households, which has implications for within household respondent selection. A 2009 survey by the US GAO of 1,143 households with at least one cell phone on satisfaction with wireless telephone services somewhat replicated those figures. However, we were able to learn more about the nature of potential sharing from analyzing additional questions on land line and cell phone use, as distinct from ownership, among adults versus children in the household. For example, among the 465 cases for which we had valid answers to the necessary questions to identify this subgroup, we found that 7% of these cell phone households reported a greater number of cell phone "users" than those who "have" (own) cell phones, which directly implies sharing. Not all of this sharing was accounted for by minors using an adult's cell phone.

Reasons for Panel Attrition: Arbitron Panelist Exit Interviews
Arianne Buckley, Arbitron Inc. (arianne.buckley@arbitron.com); Courtney Mooney, Arbitron (courtney.mooney@arbitron.com); Yelena Pens, Arbitron (yelena.pens@arbitron.com)

Arbitron has developed an electronic Portable People Meter (PPM) that can automatically detect audio exposure to encoded radio signals. We ask panelists to wear their meter everyday from rise to retire in order to measure media exposure. Arbitron conducts ongoing research to improve the sample performance of our panels.

Households are de-installed from the Arbitron panel for one of three reasons: (1) the household's time in the panel has expired, (2) the household leaves on their own because they no longer want to participate or they move and (3) Arbitron asks the household to leave as a result of poor compliance. In April 2009, Arbitron conducted phone interviews among Houston de-installed panelists to examine at how the panel experience, from recruitment to de-installation, differs between the three different de-installed groups. Completions rates were also analyzed by length of time out of the panel and by de-installation group to determine how likely panelist would be willing to complete an interview after being out of the panel for an extended period of time.

The phone interviews asked panelists who had left our panel between January 2008 and March 2009 about their overall experience in the Arbitron panel. They included general questions about the panelists' experience in the panel and questions unique to why the panelist left the panel. Analyses of these interviews looked for patterns of both similarities and differences between panelists with regards to their reasons for de-installation. Furthermore, the analyses also examined the completion rate between the three groups and how likely it is to complete an interview with those who did not have any communications with Arbitron for up to a year before being re-contacted.

Social Networks and Social Networking: Utilizing Social Networking Sites to Conduct Field Work When Your Sample Members are Likely to Know Each Other

Lisa M Setlak, NORC (setlak-lisa@norc.org)

Increased geographical mobility and the increased use of cell phones as primary numbers have made traditional methods of locating study participants more challenging. This paper explores the use of social networking sites to locate and prompt respondents to complete a half-hour web questionnaire in a multi-cohort, longitudinal, education study. The education study periodically surveyed applicants, both recipients and non-recipients, to a college scholarship program about their educational and work experiences. All applicants to the program were currently enrolled in one of sixteen high schools. Studied cohorts were surveyed one year after graduating from high school, and then in two to three year intervals. Social network methodology was used during the last two months of the last full-round of the study, during the second follow-up (third) survey for one cohort (1014 sample members). On two popular social networking sites the names and high school affiliations were used to locate sample members. Additional searches were conducted using the social networks of previously identified sample members. After being located, sample members were either sent a friend request, or a message asking if they would like to participate in the current round of the study. In cases where messages were used and there was no response to the initial contact, project staff continued sending weekly prompts. This paper will compare the percent of sample members located along with response and refusal rates for the eighth round with previous rounds in which only traditional locating and prompting methods were used. Overall, using social networking sites allowed our study to locate difficult cases and improve response rates when traditional methods had been exhausted.

Political Knowledge and Attitudes

Competitive Elections and Attitudes Toward Congress

Heather Evans, Sam Houston State University (hke002@shsu.edu)

Research on U.S. Congressional elections has placed great emphasis on the role of competitiveness. Competitive campaigns have been found to increase citizen participation, engagement, and political knowledge. However, little is known about whether competitive campaigns have lasting consequences for citizen attitudes.

Using a survey of citizens conducted one year after the 2006 Election, I investigate whether competitive House campaigns affect citizens' attitudes toward Congress. The data was collected as part of the 2007 Cooperative Congressional Election Study (CCES) and includes an oversample of respondents from competitive House races to ensure there was a sufficient number of respondents who were exposed to a competitive race. This study examines the effect of competitive elections, political interest, political activity, party identity, political ideology, interactions with legislators, and economic conditions on approval and trust in Congress.

Political Knowledge as an Independent Variable

Daniel Cassino, Fairleigh Dickinson University's PublicMind Poll (dcassino@fdu.edu); Peter J. Woolley, Fairleigh Dickinson University's PublicMind Poll (woolley@fdu.edu)

Much of modern political psychology has been built around the twin findings that (1) Americans don't know or care very much about politics and (2) their votes are fairly accurate anyway. In other words, researchers approaching the problem from the perspective of both high and low information rationality have found that individuals with low levels of knowledge about an election tend to vote about the same way that they would have if they had high levels of knowledge.

However, none of this means that there are not meaningful differences between individuals with and low levels of political knowledge. Prior research has shown that high knowledge individuals are more likely to turn out at the polls, and their opinions may more resistant to change than those of lower knowledge individuals. As such, differentiating between high and low knowledge individuals may be of value to polls.

In our study, New Jersey voters polled in three RDD likely voter samples about their views of their 2009 gubernatorial election were also asked political knowledge questions measuring both their knowledge of current events and general civics. The responses of the group with high levels of political knowledge were both more stable over time, and were much closer to the final outcome of the election. These findings suggest that, at least in off-year elections, political knowledge is far from inconsequential, and its measurement may tell us an enormous amount about the probable outcome of elections.

Measuring Televised Exposure to Politics: Evaluating a New Approach.

Susanna Dilliplane, University of Pennsylvania (sdilliplane@asc.upenn.edu); Seth Goldman, University of Pennsylvania (sgoldman@asc.upenn.edu); Diana Mutz, University of Pennsylvania (mutz@sas.upenn.edu)

Measuring media use is a perennial source of difficulty for scholars interested in political communication and public opinion. For a variety of purposes, researchers need reliable and valid measures of the extent to which people have been exposed to political content in mass media, and yet accurate assessments of this activity have proved elusive. In this study we draw on 3 waves of panel data from a representative sample of over 12,000 Americans in order to evaluate an alternative approach to measuring exposure to politics on television.

We identify three sources of problems with the measures that have traditionally been used to tap televised exposure to politics. In addition to the problems that have been documented with the reliability and validity of traditional news exposure items, we also point out how, in the changing media environment, it is difficult for even the most motivated of respondents to understand what should count as “news” or “public affairs content” for purposes of a given survey response. Traditional measures are also limited in their ability to inform researchers about the variations in content to which people have been exposed. When network news programs dominated, and content analyses suggested that they were all very similar, not knowing which program viewers watched did not seem particularly problematic. But now viewers can choose from dozens of programs that take rather different approaches to politics.

Using an unusually rich panel data set collected by Knowledge Networks for the National Annenberg Election Study, we evaluate the stability, reliability and predictive validity of television exposure measures gathered from October 2007 through February 2009. These measures employed lists of widely watched programs to try to alleviate the cognitive burden placed on respondents with traditional measures. We evaluate the strengths and limitations of this new approach based on our analyses.

The Gender Gap in Political Knowledge: Results from the National Annenberg Election Surveys 2000, 2004, and 2008.

Kate Kenski, Dept. of Communication, University of Arizona (kkenski@email.arizona.edu)

Gender gaps in political knowledge and new comprehension have been found persistently in the social science literature (Chaffee, Zhao, & Leshner, 1994; Delli Carpini & Keeter, 1996, 2000; Frazer & Macdonald, 2003; Graber, 1988; Kenski & Jamieson, 2000; Robinson & Levy, 1986). Men almost always outperform women on tests of political information. This gap is robust, having been recorded at approximately the same magnitude over the last 40 years (Delli Carpini & Keeter, 1996). Recently, however, some researchers have challenged the size and existence of the gap (McGlone, Aronson, and Kobrynowicz, 2006; Mondak and Anderson, 2004). One set of researchers argue that “approximately 50% of the gender gap is illusory” because of response sets present in traditional coding of knowledge scales. Another group of researchers argue that the gender gap in political knowledge can be explained as a function of stereotype threat felt by female respondents when interviewed by male interviewers. In this study, both of these explanations are examined and challenged using data from the 2000, 2004, and 2008 National Annenberg Election Surveys. Even when the approach recommended by Mondak and Anderson (2004) and Mondak (2001) is taken, the gender gap persists at a level much greater than the 50% illusion suggested. Moreover, the results indicate that there are some validity concerns with the Mondak approach to knowledge scale construction. When gender of interviewer is taken into consideration, the magnitude of the gap does not change. In short, results indicate the gender gap in political knowledge does exist and has significant consequences on vote choice.

The State of the Parties in the American Electorate

Michael Dimock, Pew Research Center for the People & the Press (mdimock@pewresearch.org); Robert Suls, Pew Research Center for the People & the Press (rsuls@pewresearch.org); Alec Tyson, Pew Research Center for the People & the Press (atyson@pewresearch.org); Jacob Poushter, Pew Research Center for the People & the Press (mdimock@pewresearch.org)

Political parties as public institutions are at an interesting moment in recent history. We are witnessing record levels of political polarization in the electorate, with Republicans and Democrats moving further apart across a wide range of political issues. Issues that in the past have not cleaved across party lines – such as trade, immigration and global warming – are now deeply partisan; and even many objective

evaluations, such as ratings of the national economy and one's own personal finances, are now reported through a partisan lens.

Yet at the same time, we are also witnessing record dissatisfaction with the political parties as institutions. Party leaders – both Republican and Democratic – are receiving deeply negative performance ratings, and many partisans express dissatisfaction with the values, priorities and overall direction of their own parties. The number of self-reported independents has reached at least a 60-year high, matched only by the 1992 election year in which Ross Perot drew a wide swath of Americans toward an independent identity. Today this simply appears to be a rejection of the parties themselves as no independent figurehead has arisen.

This paper will assemble data from multiple public opinion sources to document the combined trends of (A) rising polarization (B) sinking party ratings and (C) broadening rejection of the parties altogether. In doing so, it will assess the combination of forces – in particular changes in the media environment and a more politicized redistricting process – that may have contributed to the current trends. Finally, we will compare the current cycle of partisan decline to that documented in the late 1980s and early 1990s – most prominently by Martin Wattenberg and David Broder – to assess whether current trends are part of long-term cycles or whether the American political parties are at a new turning point.

AAPOR Concurrent Session A

Friday May 14, 2010

8:00 a.m. - 9:30 a.m.

Candidate Preferences & Election Outcomes

Polling a 3rd Party Challenger: Fact and Artifact

Peter J. Woolley, Fairleigh Dickinson University's PublicMind Poll (woolley@fdu.edu); Daniel Cassino, Fairleigh Dickinson University's PublicMind Poll (dcassino@fdu.edu)

How and when should the name of an independent party challenger be read by pollsters along with the names of the major party candidates?

Pollsters face a dilemma. Giving equal place to a third party or independent challenger may over-estimate his or her support by putting that candidate on an equal footing with much better known major party nominees. On the other hand, not reading the name of what might be a significant 3rd party challenger may unfairly prejudice respondents in favor of the major party candidates. Should pollsters anticipate the independent candidate's level of support by reading the independent candidate's name, or by reading the name do they create that candidate's support?

In a series of RDD likely voter studies, we employed several methods to measure New Jersey voter's preferences in New Jersey's 2009 gubernatorial election, a contest for which one independent challenger qualified for public campaign funds. While conventional wisdom suggested that the challenger's popularity quickly grew - receiving as much as 20 percent support in some polls - and eventually melted, evidence suggests otherwise: that the apparent popularity of the 3rd party challenger was little more than an artifact of question construction.

Moderators of Candidate Name Order Effects

Nuri Kim, Stanford University (nurikim@stanford.edu); Jon Krosnick, Stanford University (krosnick@stanford.edu)

The verdict has been in, for quite some time already, about the effect of ballot position on electoral outcomes: being listed first on a ballot matters. Existing work in this area had substantiated the hypothesis of primacy effect in a wide array of electoral contexts, and identified important moderators of this effect related to the characteristics of the race (e.g., salience of election, type of race), candidates (e.g., incumbency, partisan cues), and voters (e.g., education level). While most of these studies refer to Simon's (1957) satisficing principle as the theoretical mechanism underlying the effect, surprisingly little attention has been directed to the psychological conditions of the voters when making these decisions. In this paper, we present an experimental study, embedded in a large national Internet survey, that specifically manipulated the psychological state of respondents regarding their choice between two hypothetical candidates in a primary election. Respondents were designed to be either greatly or not at all ambivalent about the candidates, and we hypothesized that those who were made to be highly ambivalent toward the candidates would be more susceptible to order effects when making their choice.

We additionally implemented a manipulation that tried to reduce order effects by stating that the order of candidate names was determined randomly. The results document the expected primacy effect and identify some conditions under which the effect does not occur.

The Stability of Individual Vote Choices in Presidential Campaigns, 1952-2008.

Robert S. Erikson, Columbia University (rse14@columbia.edu); Christopher Wlezien, Temple University (wlezien@temple.edu)

Building on our past work on the stability of aggregate vote preferences, this project will examine the fluidity of individual-level vote choices during presidential campaigns. The leverage comes from panel surveys. In many cases the panel data is of two waves, one before and one after the election such as the ANES studies conducted 1952-2008. We model the connection between pre-election vote intention and post-election reports of vote choice as a function of the timing of the initial interview, measured in days before the election. One result is that the stability of vote choices varies only slightly as a function of the days before the election. This implies that the underlying latent vote preferences of voters tend to be quite stable over the campaign with only occasional short-term variation.

County Elected Officials, Public Opinion, and the Economy: The Declining Effect of Partisanship on Economic Perceptions.

Richard L Clark, Carl Vinson Institute of Government, University of Georgia (clark@cviog.uga.edu)

The 2008 Election brought about a significant change in the federal government's approach to the economic crisis brought on by problems associated with sub-prime mortgage defaults. While the changes in Washington, DC, have received a vast amount of attention, there has been little attention paid to those governing at the local level.

For six years, the University of Georgia's Carl Vinson Institute of Government, in partnership with the National Association of Counties, has polled a random sample of county elected officials on issues related to local government and the health of counties. The 2009 poll asked a series of questions about the economy and the economic crisis facing the country at the outset of 2009 and the Obama Administration. This paper will explore the relationship between county elected officials' views of the economy and their party affiliation and support for the President. In addition, this paper will contrast those relationships found between partisanship and the economy with those found in general population surveys. Taken together, the findings will document a decline in the relationship between partisanship and views of the economy.

Americans, the Two-Party System and 'Third Parties': Trends, Predictors and a Consideration of Question-Wording Artifacts.

Christian Collet, International Christian University (collet.chris@gmail.com)

The 1990s saw exceptional diversity in the American party system and one of the most significant independent presidential candidacies in history. Yet, in the 2000s, a burgeoning interest in alternatives gave way to a debate over major party polarization. Insofar as independent activity is seen as a 'response to major party failure' (Rosenstone et. al. 1996), the shift implies that the appetite for systemic electoral reform is in recession, as the country remains locked in ideological warfare. Americans, it appears, have been 'sorted' into two camps, with progressives now enjoying an advantage due to Obama, Bush/GOP unpopularity, Iraq and the economic downturn (Levendusky 2009).

What polarization discourse has overlooked is not merely the stability in social attitudes (Fiorina et. al. 2004, 2008), but Americans' ongoing ambivalence toward the major parties (e.g., NBC/WSJ, Oct 2009;

Harris, Jul 2007; CBS, Jun 2007; Post/Kaiser, May 2007.) As stimulus and health care debates have engulfed partisan Washington, the proportion identifying as Independents has reached, according to Pew, historic levels (“Independents...”, May 21, 2009). The majority saying they support a ‘new party to compete’ with the Democrats and Republicans has not changed since 1996 (CBS/NYT).

The question of ongoing independence – and the meaning of Americans’ desire for alternatives – are the focus of this paper. It takes on three tasks. First, it updates previous work by extending an analysis of trends toward the two party system and third/alternative parties since 1938 (Collet 1996). Second, it considers the potential effects of question wording on expressions of major party discontent and alternative party support. Third, using data from 1994/2007 Gallup and 1996/2007 CBS polls, it provides the results of multivariate analysis to help account why Americans profess a desire for party alternatives – results that may discern whether a resurgence in independent activity is viable.

Considerations in Choosing Between the American Community Survey & the Current Population Survey

Explaining Differences in Survey Estimates of Health Insurance Purchased Directly by Consumers.

Victoria Lynch, Health Policy Center, Urban Institute (vlynch@urban.org); Michel Boudreaux, University of Minnesota (boudr019@umn.edu)

Estimates of public health insurance coverage and uninsurance often vary between surveys and this phenomenon has been well studied. Less attention has been paid to divergent estimates of insurance purchased directly by consumers. The American Community Survey (ACS) found that 14.2% of the population had direct purchase coverage on an average day in 2008, and the Current Population Survey (CPS) found that 8.9% had direct purchase coverage for at least one day that year. To make sense of these differences this study considers survey design characteristics that may influence reports of direct purchase. We propose that the ACS’s mode of administration and the minimal nature of its health insurance question lead to misclassification of public plan enrollees into direct purchase. Surveys like the CPS, which are predominantly administered by CATI/CAPI, use state-specific fills for the names of insurance programs, and have coverage questions aimed specifically at children may do a better job of prompting respondents to correctly report public coverage.

To evaluate the possible influence of mode we compare demographically adjusted direct purchase rates from each mode of the ACS to results from other surveys. To evaluate possible misreporting, we use regression models to determine what proportion of the population with direct purchase is likely eligible for public coverage.

Directly purchased coverage will increasingly become an important coverage category to measure as policymakers seek to expand access to health insurance by expanding individual market regulations and by subsidizing low-income consumers to purchase coverage on insurance exchanges. The ACS will be an important tool for monitoring direct purchase coverage and this analysis will help inform data users on the strengths and weaknesses of its health insurance item. Furthermore, our analysis will benefit survey methodologists designing their own surveys.

Measuring High School Completion: Comparisons Between the American Community Survey and the Current Population Survey

Angelina N. KewalRamani, American Institutes for Research (akewalramani@air.org); Lauren Frohlich, American Institutes for Research (lfrohlich@air.org)

Over the past four decades, the Current Population Survey (CPS) October supplement has provided national-level estimates on education, specifically school enrollment and educational attainment. Since 2005, the American Community Survey (ACS) has also collected information on enrollment and attainment in education. The two surveys differ in several aspects, including sample size, mode of administration, questionnaire design, and reference period. For instance, the October CPS includes a battery of detailed questions on enrollment in school or college and the highest level of education completed. In comparison, the ACS includes three questions on these topics. Unlike the CPS, the ACS includes a rolling reference period which means that some respondents will have completed one less year of school than if they were interviewed in the October CPS. While the October CPS is a rich source of data, it has some limitations when compared to the ACS. Due to the large sample size of the ACS, it allows for more detailed comparisons of education measures by state and other demographic characteristics, including detailed racial/ethnic subgroups. This paper will examine methodological differences between the ACS and October CPS in measures of high school completion. In addition, ACS and CPS estimates of high school completion by different demographic characteristics will be compared for 2007 and 2008. The analysis will focus on ACS estimates restricted to the civilian, non-institutionalized population which is the population covered by the CPS. The research will show whether ACS estimates of high school completion are comparable to the CPS, and whether the ACS can supplement the long-term trend estimates provided by the CPS.

Health Insurance Estimates from the CPS versus the ACS: the Role of State-Specific Program Names.

Joanne Pascale, U.S. Census Bureau (joanne.pascale@census.gov)

Since the late 1980s the U.S. Census Bureau has been producing estimates of health insurance coverage based on its Current Population Survey (CPS). However, beginning in the fall of 2009 the Census Bureau began releasing health insurance estimates from the American Community Survey (ACS), which has a substantially larger sample size and thus enables more detailed analysis at the state-level than the CPS. Researchers are now examining how the two surveys compare, in terms of methodology and estimates, and one area of concern is the role of state-specific names in questions about coverage through government health programs such as Medicaid and the Children's Health Insurance Program (CHIP). The CPS makes an effort to embed all of these state-specific names into the questionnaire, while the ACS does not (due to the complexities of producing state-level forms and the space constraints on the ACS mail-out/mail-back form). In some states the state-level program name is fairly straightforward (e.g., "Medi-Cal" in California) but in other states the program names are less obvious (e.g., "DC Healthy Families" in Washington, D.C., and "SoonerCare" in Oklahoma). Furthermore, in some states there are upwards of nine or ten names of programs offered through the state. Thus there is considerable variation in both the nature of state-level program names and the sheer number of names. There have been some isolated state-level studies on the effects of embedding the state-level program name as a stimulus in the questionnaire, but very little is known about these effects across states in general, considering the variation in naming conventions. This paper will examine the relationship between the nature and number of state-specific program names and levels of reporting in both the ACS and the CPS across states.

Income and Health Insurance Coverage: Comparing the American Community Survey and the Current Population Survey.

John L. Czajka, Mathematica Policy Research (jczejka@mathematica-mpr.com)

In 2008, health insurance coverage was added to the information collected in the American Community Survey (ACS). With this change, health policy analysts now have a data source that can provide reliable estimates of health insurance coverage below the state level as well as more precise state-level estimates than the Current Population Survey (CPS). Because health insurance coverage is strongly associated with family income, it is important that prospective users understand how estimates of income may differ between the ACS and CPS. Earlier research found surprisingly strong similarities between the two surveys' estimates of the distribution of income despite substantially different levels of detail in their income questions, but there were notable differences as well. Furthermore, users need to be aware of conceptual and measurement differences regarding reference periods and the treatment of college students and persons who are unrelated to the householder. This paper extends the earlier research to include comparative estimates of health insurance coverage by income level and addresses a number of issues in using the ACS for health policy analysis.

Mode of Administration and Other Design Considerations in Health Insurance Estimation

Jeanette Ziegenfuss, Division of Health Care Policy and Research, Mayo Clinic

(ziegenfuss.jeanette@mayo.edu); Michel Boudreaux, University of Minnesota

(mhboudreaux@gmail.com)

In the fall of 2009, the Census Bureau released health insurance coverage estimates from the American Community Survey (ACS) for the first time. Overall estimates of uninsurance are very similar between the ACS and the Current Population Survey's Annual Social and Economic Supplement (CPS) (15.1 percent uninsured in the ACS compared to 15.4 percent in the CPS). This similarity is somewhat surprising given the substantial differences in data collection methods, questionnaire design, and conceptual definition of insurance across these two surveys. Of note, approximately half of ACS surveys are completed via mail whereas the CPS does not have a mail component. Further the CPS asks about health insurance coverage for the entire family at once while the ACS asks about coverage individually for each family member. And finally, the CPS contains multiple questions about different types of health insurance coverage followed by a verification question whereas the ACS has only one question stem per person that is followed by a list of coverage types in a forced-choice format. With respect to conceptual definition of coverage, the ACS asks about current coverage whereas the CPS is intended to measure coverage over the entire prior calendar year. By looking at difference of reported health insurance coverage by mode, survey, and respondent (self compared to proxy-report) we provide evidence that the surveys are not performing as similarly as the overall rate of uninsurance would suggest. Understanding the viability of the ACS estimates is important to policy makers and survey methodologists alike. Findings can be used by entities such as states interested in measuring health insurance coverage by mail in light of costs associated with RDD surveys and potential cell phone coverage bias.

Household Telephone Access: Coverage & Adjustment

Arbitron's Hybrid Sampling Frame Approach Balances Coverage and Response

Alan Tupek, Arbitron Inc. (alan.tupek@arbitron.com); Robin Gentry, Arbitron Inc.

(robin.gentry@arbitron.com)

Arbitron now uses a hybrid sampling frame approach in all markets it surveys. An RDD landline sampling frame is used to cover households with one or more landline telephone, and an address frame is used to identify and cover cell phone only (CPO) households. The hybrid approach allows Arbitron to better reach households for which the address and telephone number can't be matched by its sampling frame

suppliers. The hybrid approach helps mitigate address frame coverage problems that include the lag time of acquiring information from the US Postal Service, communities and buildings with a single drop point address, and a significant, but declining number of rural route post office boxes. In addition, we find that over ten percent of addresses are undeliverable. It is also difficult to obtain a high cooperation rate from those households on the address frame without phone numbers. Based on coverage studies that Arbitron has conducted, the advantages of its hybrid sampling frame approach are discussed.

Cell-Phone Sampling: An Alternative Approach

David B. Lambert, Research consultant (dblambert@verizon.net); Gary Langer, ABC News (gary.langer@abc.com)

Most current approaches to cell-phone sampling consist of mixed-mode designs that require non-empirical decisions in matters such as the proportion of cell vs. landline sample released, the proportion of cell interviews completed and the weights applied to cell-only, cell-plus-landline and landline-only samples. An alternative approach is to devise a dual-mode sampling plan that incorporates cell-only respondents as a specific issue of known noncoverage. This approach benefits from straightforward theoretical and empirical underpinnings, obviating guesswork in sampling and weighting; what it lacks is the inclusion of “cell-mostly” respondents. Our paper, drawing on more than 25,000 total interviews in ABC News/Washington Post polls since fall 2008, describes the theoretical and empirical bases for this approach, its operationalization in sampling, sample stratification and weighting, its cost efficiency and its impact on demographic and other estimates. We include comparisons to other cell/landline sample designs and results. Presentation by David Lambert, research consultant; and Gary Langer, director of polling, ABC News.

Household Telephone Configuration

Karol Krotki, RTI International (kkrotki@rti.org); Andrew Peytchev, RTI International (apeytchev@rti.org)

As survey research adapts to the increasing prevalence of cell phones, research is needed on cell phone use within households, especially households that have both landline and cell phone service. Researchers have studied problems associated with the dual frame approach but little is known about households that make up the intersection of the two frames, the numbers of different telephone numbers, both landline and cell phone, linked to such household, and the distribution of respondent characteristics related to those numbers. This information is needed not only to accurately calculate selection probabilities but also to establish efficient sampling and interviewing procedures.

In this research we interviewed a sample of close to 1,000 cell phone owners nationwide and collected information on the number and type of phones in the households as well as the sharing patterns among household members. The results are informative of the different configuration patterns to be found in households, shed light on telephone ownership and usage patterns, and suggest strategies for how to efficiently design telephone surveys in the cell phone era.

Measuring the Number of Land Line and Cellular Telephones Used for Voice Calls in Households to Properly Weight RDD Surveys for Unequal Probability of Selection

Ana Villar, Stanford University (villar@stanford.edu); Jon A Krosnick, Stanford University (krosnick@stanford.edu); Matthew DeBell, Stanford University (debell@stanford.edu)

It is well known in survey research that data collected in RDD telephone surveys must be weighted to reflect the fact that some households have a greater probability than others to be in a sample, due to the

fact that multiple land lines or cell phones can reach the household for voice calls. Therefore, RDD telephone surveys routinely include question sequences designed to measure the number of land and cell phone lines that could reach a household. Surveys administered by the National Center for Health Statistics face to face to national representative samples of American Households have used such question sequences to track the increases in the number of cell-only households in recent years. This paper began with a review of existing question sequences and identified shortcomings of those sequences that might cause measurement error in assessing the number of land lines and cell phones that can reach a household for voice calls. Therefore, the American National Election Studies conducted an experiment by designing and administering a new set of questions for this purpose, designed to yield more accurate measurements. The question sequence was also administered in the Face to Face Recruited Internet Survey Platform (the FFRISP). This paper will report an evaluation of the success of the question sequence in improving measurement accuracy.

Media & Public Trust

Media Channel Effectiveness and Trust

Edward Paul Johnson, Western Wats (ejohnson@westernwats.com); Dan Williams, Western Wats (dwilliams@westernwats.com)

Campaign messaging and pollsters are facing similar and corresponding challenges in reaching, identifying and segmenting diverse groups of voters. The fast moving world of electronic communications creates rapidly shifting waves of messaging options. The 2008 campaign taught many the value of messaging in many modes and in identifying to whom the messages should be sent. Will people trust a media report via the radio, TV or newspaper or a trust a social media message from a friend? In both messaging and polling, an issue of trust in media mode will be important in determining winners and losers in future elections.

This research aims to assess current levels of participation and trust in different media sources in research conducted via traditional RDD phone, online access panels and members of a social media community. We start by identifying the most trusted media source from TV news, newspapers, radio, and news websites. The winner of this source will represent the high professional, but low level of traditional news media. Then we identify the most trusted social media from You-Tube, Facebook, Twitter/texting, and web blogs to compare with the traditional media source. Two volatile issues (health care reform and Obama's presidential performance) are then addressed. Respondents are given two conflicting messages in a random order, one from the social media and the other from the traditional media. Lastly we gauge whether and in what direction their opinion shifts on the topic.

AAPOR areas and issues addressed or potentially addressed:

- Surveying and interviewing diverse populations
- Media influence on public opinion
- Internet and other web-based data collection

Facilitating Expressive Participation via Trust in Media. Evidence From An Experiment on an Environmental Issue.

Eulalia Puig Abril, UW-Madison (epabril@wisc.edu); Jill E. Hopke, UW-Madison (jehopke@wisc.edu); Albert Gunther, UW-Madison (agunther@wisc.edu); Hernando Rojas, UW-Madison (hrojas@wisc.edu)

One of the critical antecedents of participation, trust, has been declining in the United States over the last few decades (Putnam, 1995)—including trust in media. Even though research has shown how certain

media such as newspapers can bolster trust, newspaper readership is on the decline (Putnam, 2000). Moreover, little inquiry has been done regarding how to increase trust at the society level, or at the level of content production. Trust becomes even more important when it comes to participation in topics that are controversial to society. This study contends that (a) there are ways in which journalists can affect trust without extending additional demands on news reporting; and (b) there are ways in which society can achieve higher levels of trust in media through instruction about journalistic standards. We employed an online experiment (n=129) conducted during 2009 among environmental groups that either favored or opposed biofuels, as well as university students from environmental majors.

To assess individual-level media effects, we tested for the effect of number of arguments in a news story, employing a news story with either two or four balanced arguments about biofuels. To assess for the society-level media effects, we tested for the effect of a media literacy intervention, with a group either receiving a media literacy video before the news story, or just the news story.

Results depict a strong, positive path from both experimental treatments to trust in media, and from there to expressive participation. Trust in media is increased with exposure to more arguments or media literacy education. In turn, this relationship leads to more expressive participation. The direct effect of either manipulation to expressive participation was not significant. Our findings suggest that both more in-depth coverage and society-level interventions in the form of media literacy could increase expressive participation via trust in media.

The hostile media phenomena and the quality of journalistic practices: An experiment regarding the debate over biofuel implementation

Michael A. Cacciatore, UW-Madison (mcacciatore@wisc.edu); Vidal Quevedo, UW-Madison (viqual@gmail.com); Hernando Rojas, UW-Madison (hrojas@wisc.edu); Albert Gunther, UW-Madison (agunther@wisc.edu)

Assessments of news have reached their lowest levels in decades. The Pew Research Center estimated in 2009 that only 29% of Americans trust media to “get the facts straight,” and 18% think it “deals fairly with all sides.” While part of the frustration with media stems from a reconstitution of the media environment in a new technological context, another component can be traced back not to the news itself, but rather to its biased processing. This later view, the hostile media effect (HME), suggests that people, especially those with strong issue positions, have a tendency to perceive news as biased against their point of view. Studies explaining the mechanisms underlying HME have pointed out that selective categorization (classifying content differently) and differing standards (perceiving a different “neutral” point) are central. Thus, it could be that a “better” news story (i.e. a more in-depth story that captures more arguments in favor and against an issue), could actually result in increased perceptions of bias.

In order to test this proposition we conducted an experiment that manipulated the number of arguments given in a news report (biofuel implementation). Participants were recruited from groups highly involved in the debate over biofuels. Subjects, randomly assigned to one of two conditions, received an article that included two arguments in favor and two against biofuels (less arguments) or an article that contained two additional arguments in favor and against (more arguments). Both articles maintained comparable argument strength. After exposure to the articles, subjects answered questions assessing story credibility, hostility, trust in journalists to cover this issue, and general media trust. Results indicate that subjects in the condition with more arguments perceived greater story credibility, issue trust, and media trust. The implication of our findings for journalistic routines and public opinion formation are discussed

When Fiction Affects Reality: Examining the Influence of Fictional Political Film on Public Opinion about Voter Fraud.

Heather LaMarre, University of Minnesota (hlarre@umn.edu)

Recent studies have begun examining the effects of political entertainment on individual-level socio-political attitudes related to a variety of public policy issues (e.g., women's rights, Holbert, Shah, & Kwak, 2003; capital punishment and gun ownership, Holbert, Shah, & Kwak, 2004). Gamson and Modigliani (1989) found that viewing *The China Syndrome* led to significantly less public support for and more negative opinions of nuclear power plants. This study provides a framework for understanding the potential role blockbuster films and Hollywood productions play in influencing political attitudes and policy support for public issues.

Building upon this earlier research, two experiments were conducted to examine the influence of political entertainment films on public opinion. Data from the experiments (study one N = 302; study two N = 198) were used to assess the direct and indirect effects of major motion picture films (*Man of the Year* and *Sicko*) on attitudes and opinions about voter fraud and nationalized healthcare, respectively.

Using structural equation modeling, the study one model revealed that viewing a fictional movie about electronic voting fraud (*Man of the Year*) led to increased negative elaborations and opinions about electronic voting in national elections. Additionally, the group who watched the film reported significantly more concerns, worries, and fears about voter fraud in the 2008 Presidential election. This process was partially mediated by how engaged audiences were in the film.

Similarly, study two results also revealed a significant relationship between film viewing and opinion about current political issues. Using MANCOVA, study two revealed that exposure to *Sicko* was significantly related to decreased support for nationalized healthcare. Additionally, after watching *Sicko*, participants reported significantly less trust in private insurance companies and significantly less favorable attitudes toward healthcare and medical supply companies than those who did not watch the film.

Remarkably Stable Public Opinion in a Turbulent Policy Environment

Ward Kay, George Mason University (wkay1@gmu.edu)

A common theme amongst critics of public opinion is that on many issues that public doesn't have an opinion. This theme is supported by survey researchers from Converse to Zaller to Bishop who have described non-attitudes or pseudo-attitudes of survey respondents to policy questions, sometimes randomly, sometimes based on cues in the question or questionnaire itself. The theory is that attitudes on non-salient issues are random and will change if the issue becomes more salient. This study contradicts that viewpoint by showing the public opinion on immigration issues has remained fairly consistent through two different focusing events that raised the salience of the issue. In spring 2006, demonstrations were held around the country by immigrants to raise awareness of the importance of immigrants in the country and to prevent a harsh immigration bill from becoming law. In spring 2007, after the switch in party-control in Congress, a second immigration bill was being considered which was opposed by the right-wing media and again salience was increased. Media polls using the same question-wording before, during, and after these focusing events has shown remarkably consistent public opinion.

Methodological Briefs: Improving Measurement & Reducing Error

Political Cocooning: Explaining How Voters Are Receptive to, Reassured by, and Rationalize Political Information on the Economy. Michael Brogan, Rider University (mbrogan@rider.edu)

The economic vote is based on an interactive relationship where voters' economic perceptions are conditioned by their partisanship and political knowledge. Voters' economic perceptions are not defined solely by economic indicators; rather, they are also influenced by the process of political cocooning. This process ensures that voters typically exhibit the following behaviors when forming their economic

perceptions: Voters are receptive to sources of information that conform to their own beliefs, reassured by sources of information that support their existing preferences, and inclined to rationalize information that reinforces their own firmly held political views. These behaviors among voters complicate the economic voting process. The degree to which one is likely to vote based on economic perceptions is either diminished or heightened based on one's level of partisanship and political knowledge. If one's partisanship is strong, then economic perceptions are not as influential in vote choice. If one's political knowledge is strong, one's partisanship is often reinforced, limiting the power of economic perceptions in vote choice. Note that those who are most informed tend to be more partisan than the average voter. The informed voter generally searches for information that is focused on her party preferences; this, therefore, influences how she views the economy. This phenomenon helps explain why economic voting tends to be asymmetric—namely that incumbents are punished more for economic downturns than rewarded for periods of prosperity (Lewis-Beck and Stegmaier 2007), as well as why voters tend to vote against their economic interests (Popkin 1992; Zaller 1992; Bartels 1996, 2002; Evans and Anderson 2006). The political cocooning model will disentangle the moderated relationships among these variables (i.e., economic perceptions, partisanship and political knowledge) by estimating and testing nonlinear marginal effects and their standard errors. The data used to test the model will come from the ANES and BES studies.

Availability of Driver's License Master-Lists For Use in Government Sponsored Research.
Matthew Walsh, University of Wisconsin-Madison (walsh2@wisc.edu); Amy Trentham-Dietz, University of Wisconsin-Madison (trentham@wisc.edu); Nora Cate Schaeffer, University of Wisconsin - Madison (schaeffe@ssc.wisc.edu); Mari Palta, University of Wisconsin - Madison (mpalta@facstaff.wisc.edu)

Although the percentage of drivers with valid licenses varies from state to state, it has historically been high enough to constitute a useful sampling frame for many public health purposes. The largest coverage problems have resulted from individuals who are not old enough to drive and those who have stopped renewing or never obtained their driver's licenses. Over the past decade states have had to restrict access to this information in order to comply with the Drivers Privacy Protection Act (18 U.S.C. 2721-2725). We are conducting a survey of all 50 states and the District of Columbia on the availability of master-lists of licensed drivers to be used to contact citizens of each state for research purposes. For each state, a hypothetical situation requiring driver's license data (name, street address, city, and zip code) was sent to the responsible government agency for review. The hypothetical situation assumes that a state university or state department of health agency is requesting the data to contact individuals for participation in a breast cancer research study. In addition, we are collecting data on opt-out mechanisms available to drivers, the costs to researchers, and additional state privacy policies that restrict use of this information. Data collection will be completed by January 2010. We currently have completed surveys from 25 states with 8 states (32%) agreeing to provide master files of licensed drivers to be used to contact individuals. State level data on the codification of privacy laws, specific rulings on the hypothetical request, and coverage issues with specific states will be presented. Results from this survey should inform researchers of the feasibility of using this sampling frame in practice.

Gender-of-Interviewer Effects in the Women's Nation Poll
Kelly Daley, Abt SRBI (k.daley@srbi.com); Allison Ackermann, Abt SRBI (ackermann@srbi.com)

Despite a wealth of literature on interviewer effects, few recent studies have re-visited gender-of-interviewer effects on questions related to gender roles and policy: published literature looking specifically at telephone survey response to questions regarding the women's movement, inter-personal gender

relations and gender policy dates to the 1990s, and the results are mixed (Huddy et al, 1997, Kane and Macaulay 1993).

Using data from the 2009 Women's Nation Survey, this paper examines gender-of-interviewer effects on response to survey questions related to gender roles and policy. The survey, conducted by Abt SRBI, included 3413 20-minute interviews with adults drawn from a national RDD probability sample of U.S. households and a cell-phone supplement. The RDD component includes oversamples of African Americans and Hispanics and interviews were conducted in English and Spanish. Male and female interviewers were randomly assigned to conduct the interviews. Interviews were conducted in August, 2009. The survey probed opinions about the tremendous influx of American women into the workforce; its impact on the beliefs and behaviors of men and women at work and home as well as government and private sector policy on the workplace and family. Equal numbers of male and female interviewers conducted interviews with randomly selected adult household members.

In this paper we test for the presence of gender-of-interviewer effects on gender role attitudes in three topic areas related to gender equality: social change, gender roles in the workplace and interpersonal relationships: specifically, do men and women differ in their responses on gender roles when asked their opinion about societal change versus in their interpersonal relationships and does gender-of-interviewer affect these responses? Next we will examine which individual or combinations of demographic characteristics increase are more likely to increase the gender-of-interviewer effect.

Public opinion polls and estimates in some European countries.

Non-coverage bias in telephone surveys. Volker Hüfken, University of Duesseldorf, Department of Social Sciences (huefkenv@phil-fak.uni-duesseldorf.de)

Currently, only a few publications of politically related surveys examine the potential of non-coverage bias (Keeter, 2006; Brick, Brick, Dipko, Presser, Tucker, & Yuan, 2007; Pew, 2008). Brick et al. (2007) point out that "the cell respondents were less likely to say they read the editorials on most days" and among voting two variables shows no significant differences between phone status. Keeter (2006) reports that cell phone only "CPO voters thus far appear to be only slightly more liberal or Democratic in their orientation". During the 2008 presidential election, Pew (2008) observed similar patterns. Drawing from the results this study seeks to replicate similar objectives.

Data were obtained from the European Social Survey (ESS4) for 2008/9. The ESS4 (first release) includes 33,264 cases. The regression models predicted 'nonvoting', 'disinterested in politics', 'left political orientation', 'were not reading newspapers about politics and current affairs' and 'were not watching news or programs about politics and current affairs'. The reported ratios of odds were adjusted to account for group differences in gender, age, education, employment status, household income, occupational class, household size, being born in the country, and subjective urbanization.

Exploring Sources of Error in the Consumer Expenditure Survey - Results from a Small-scale Validation Study

Scott Fricker, Bureau of Labor Statistics (fricker.scott@bls.gov); Jennifer Edgar, Bureau of Labor Statistics (edgar.jennifer@bls.gov)

This study examined how data quality and respondents' perceptions of the survey-taking experience were affected by aspects of the survey design (i.e., mode of administration), respondent behaviors during the survey (i.e., use of records, use of recall-aid flashcards), and respondent characteristics (e.g., household structure, expenditure patterns, record-keeping strategies, etc.). Respondents in this study were administered sections of the Consumer Expenditure Quarterly Interview Survey (CEQ) in one of two

randomly assigned modes of administration - telephone or personal visit. In addition to capturing respondents' expenditure reports, we recorded their use of records and survey recall aids, indicators of fatigue or frustration, and other qualitative data throughout the interview session. Upon completing the CEQ questionnaire, respondents were then administered a semi-structured debriefing interview that probed their survey experience (e.g., perceptions of burden, reliance on records, effectiveness of the CEQ recall aids) and the impact of survey mode on data quality and respondent attitudes. We examined the types of records that respondents used in survey, attempted to validate respondents' expenditure reports, and examined when and why reports deviate from actual expenditures.

Differences in Length of Survey Administration between Spanish-Language and English-Language Survey Respondents

Tom Wells, Knowledge Networks (twells@knowledgegenetworks.com); Monica Vidalon, Knowledge Networks (mvidalon@knowledgegenetworks.com); Charles DiSogra, Knowledge Networks (cdisogra@knowledgegenetworks.com)

In 2008, Knowledge Networks expanded KnowledgePanel to create KnowledgePanel Latino, the first nationally representative online sample of Latinos in the U.S. We have observed that the Spanish-language panelists take more time to self-administer their Spanish-language online questionnaires than those taking the English version – both Latinos and non-Latinos range 50% to 100% longer. For example, in our 2009 demographic profile survey, median completion length is 8 minutes among roughly 40,000 English-language panelists, while 21 minutes among roughly 2,000 Spanish-language panelists. This pattern holds for different survey topics and has practical implications for respondent burden, respondent incentives and costs.

To our knowledge, this is a new finding, as there are no published research findings on these differences in online surveys.

We will evaluate several possible explanations for this finding: longer Spanish questionnaires in terms of word count; skip patterns, respondent demographics (e.g., education, age, household size), cultural differences, panel tenure, Internet connection speed, and survey-taking behavior.

We employ regression to estimate the effects of survey language, demographics, panel tenure, and type of Internet connection on survey completion time.

Our regression results show that education, age, and broadband internet connection have large effects and significantly reduce survey completion time. However, once several factors have been controlled, Spanish-language survey takers still take, on average, 7 minutes longer than their English-language counterparts to complete the main demographic profile survey. Spanish survey language still remains a strong predictor of administration length, which may speak to cultural differences and survey taking behaviors, which we will further investigate in this paper.

Explaining 'House Effects': Describing and explaining differences between Overt Racism measures in GSS and ANES.

Gaurav Sood, Stanford (gsood@stanford.edu)

The General Social Survey (GSS) and American National Election Studies (ANES) are two pre-eminent surveys tracking over-time trends in the US adult population in terms social and political attitudes, beliefs and behavior. Very few questions are asked identically in the two surveys, but some questions measuring racial attitudes have been asked in many years in both surveys, allowing a comparison of distributions

and trends. If the GSS and ANES track trends as expected, then they should produce very similar results. Systematic differences are a cause for concern. Utilizing data from the 1996, 2000, 2004, and 2008 GSS and ANES, we explored the distribution of overt racism measures across both polls, and over time trends. We find that in the overall sample, and when limiting the sample to just whites, mean overt racism ratings are always more positive in the ANES, though not always significantly, than GSS. We also find that people are far less likely to choose the midpoint in ANES than the GSS. To explain these differences, we look to time of administration effects – overt racism measures are offered in the post-election interview in the ANES while the questions are offered in period leading up to election in the GSS – and find some evidence that intervening election is indeed part of the explanation using other nationally representative surveys asking similar questions before elections.

Non-response, Paradata, & Data Quality

A Comparison of Alternative Quality Measures for Survey Data

James Wagner, University of Michigan/Survey Research Center (jameswag@isr.umich.edu)

There is a growing recognition that the response rate is an incomplete measure of the quality of survey data relative to the risk of nonresponse bias. We know from theory that nonresponse bias is a product of both the nonresponse rate and the differences between responders and nonresponders. Theory allows for the possibility that a higher response rate could produce the same or worse nonresponse biases. Recent empirical studies have confirmed that the response rate may be a poor predictor of bias. As a result, the search is on for alternative measures that may help us evaluate the risk of nonresponse bias in real-world situations.

This paper will focus on a comparison of a set of proposed alternative measures of the quality of survey data in relation to nonresponse bias. The strengths and weaknesses of each will be discussed. Where possible, relationships among these measures will also be discussed. Real-world applications of each measure will be introduced to aid in this comparison.

Finally, it will be argued that since nonresponse bias is such a complex, multi-dimensional problem, a multi-faceted approach to monitoring for the risk of this bias will also be needed. Each of these measures may be an appropriate part of a toolkit survey researchers need as they evaluate their data.

Classifying Eligibility

David Chearo, The University of Chicago Survey Lab (dchearo@uchicago.edu)

Properly determining the eligibility of cases for participation in a research project is vital to the representativeness paradigm, in which collected data from a sample of a population are thought to represent characteristics of the larger population. Determinations of eligibility also shape rates of response, cooperation and refusal, all of which are often seen as indicative of the quality of data collection methods and the representativeness of collected data.

AAPOR's Standard Definitions provide a framework for classifying eligibility into four broad categories: interviews, eligible non-interviews, unknown eligible cases, and ineligible cases. These categories can be universally applied, but elaboration on the Standard Definitions is necessary to show how the categories work in general and in research designs using multiple modes and frames.

In this paper I build upon the Standard Definitions to work toward a more refined general description of eligibility and the way it should work. I assert that all cases begin in a state of unknown eligibility, in part due to inherent and often immeasurable sampling error. From there, eligibility can only be determined through a limited set of transitional instances that can be generalized. I provide taxonomy of such instances and demonstrate the mechanics through which cases come to be eligible or ineligible. Additionally, I discuss considerations for different combinations of frame and mode and make recommendations for designing research to properly determine and document eligibility.

Who is Responsible for the Bias? Using Classification Trees to Identify Subgroups of Likely Nonrespondents and Assessing their Relationship to Key Survey Estimates Using Structural Equation Modeling.

**Morgan S Earp, USDA National Agricultural Statistics Service (Morgan_Earp@nass.usda.gov);
Jaki S McCarthy, USDA National Agricultural Statistics Service (jaki_mccarthy@nass.usda.gov)**

With increasing nonresponse rates, the bias of estimates is of growing concern. The USDA National Agricultural Statistics Service (NASS) conducts the annual Agricultural Resource Management Survey (ARMS). ARMS collects detailed economic data from a sample of US agricultural operations and suffers from relatively low response rates for a federal survey. Currently NASS' research on ARMS nonresponse assesses characteristics of nonrespondents and nonresponse bias separately. Using classification trees, NASS has identified characteristics associated with nonresponse and thus subgroups in the population worth targeting to improve response rates; Using Census of Agriculture data as a proxy of ARMS data, NASS has assessed the nonresponse bias of key estimates. However, currently it is unknown whether all subgroups of likely nonrespondents contribute to the nonresponse bias of key survey estimates equally. It is possible that some subgroups contribute more to the bias of some estimates than others do, and that some subgroups may be likely nonrespondents, but are not contributing to any bias in the estimates.

This paper discusses the use of classification trees and structural equation modeling (SEM) to determine the relationship between response propensity and nonresponse bias. Unlike research done by other agencies using a similar approach, NASS possesses matching Census of Agriculture data at the record level, rather than information at the zip code or county level; Therefore, NASS is capable of making proxy comparisons to determine which estimates are most heavily influenced by various subgroup response propensities. Classification trees were used to arrange ARMS sample units into subgroups with like response propensity. SEM was used to assess the relationship between subgroup response propensity scores and key survey estimates.

This research will enable NASS to flag likely influential nonrespondents and target data collection efforts to minimize bias in survey estimates.

Examining Paradata of Minority-Owned Businesses in the Kauffman Firm Survey

Betsy Santos, Mathematica Policy Research (bsantos@mathematica-mpr.com); David DesRoches, Mathematica Policy Research (ddesroches@mathematica-mpr.com)

Due to their underrepresentation in the U.S. economy, minority-owned businesses continue to be of significant interest to researchers studying entrepreneurship, innovation, and business productivity. The results of several national surveys have indicated that African Americans and Latinos have lower rates of self-employment and business ownership (Robb and Fairlie, 2007). Also minority-owned businesses in some industries have more difficulty obtaining financing than white-owned businesses (Grown and Bates, 1991). In this context, the need to maintain sufficient minority-owned businesses in longitudinal surveys of

new businesses takes on additional importance, and understanding the behavior of business owners asked to participate in surveys becomes essential to researchers who want to study this population.

This paper will analyze paradata from the baseline and four follow-up surveys of the Kauffman Firm Survey (KFS), a longitudinal survey of new businesses in the U.S., to determine what, if any, differences exist between minority-owned businesses (defined as being more than 50 percent owned by one or more owners who are Hispanic or African-American) and their white counterparts in the data collection process. Specifically, the paper will examine response rates, timeliness of response, cooperation and refusal rates in each round as well as longitudinally. The paper will also examine differences in mode of interview, and data quality measured through item nonresponse to key survey questions.

Provider Nonresponse Error Assessment in the National Immunization Survey

Elizabeth Ormson, NORC (ormson-elizabeth@norc.org); Kennon R Copeland, NORC (copeland-kennon@norc.org); Kirk M Wolter, NORC at University of Chicago (wolter-kirk@norc.org); Philip J Smith, National Center for Immunization and Respiratory Diseases (CDC) (pzs6@cdc.gov)

Survey nonresponse reduces effective sample sizes, thereby increasing variance of survey estimates, but also may yield bias in the estimates. Nonresponse error has become an even greater concern as response rates for surveys have been declining in the past decades. The magnitude of non-response bias is a function of response rate and the difference between respondents and nonrespondents with regard to the variable of interest. Recent work has suggested increasing nonresponse rates may be less directly linked to increased nonresponse bias than are differences between respondents and nonrespondents. In many surveys, it is difficult to estimate nonresponse error as data to compare respondents and non-respondents are not available.

This paper examines nonresponse error in the National Immunization Survey (NIS) – a nationwide, list-assisted random-digit-dial (RDD) survey sponsored by the Centers for Disease Control and Prevention and fielded by NORC – that monitors the vaccination rates of children aged 19-35 months and adolescents aged 13-17 years. Data collection for the NIS involves both a household interview and a provider survey, each of which is subject to nonresponse error. One objective of the NIS-Registry project was designed to examine potential coverage, response, and provider nonresponse error in the NIS. A sample of NIS-eligible children was selected from the state immunization information system (IIS) in Arizona and Michigan. The NIS household and provider-record-check (PRC) questionnaires were administered for both samples during Q1/2008 and Q2/2008.

Using registry data, comparisons between sample cases with adequate provider data and sample cases with missing adequate provider data allow assessment of the nature and impact of this portion of NIS nonresponse. Sociodemographic characteristics and vaccination coverage rates based on the registry data for respondents and nonrespondents are compared to assess the impact of nonresponse error in the NIS.

Public Attitudes Towards Climate Change & Energy Use

Effects of Priming and Extremity of Response on Individual Support for Alternative Energy: Evidence from a Survey-Based Experiment.

Andrew R. Binder, University of Wisconsin - Madison (arbinder@wisc.edu); Michael A. Cacciato, University of Wisconsin-Madison (mcacciato@wisc.edu); Nicholas Brigham Schmuhl, University of Wisconsin-Madison (nbschmuhl@wisc.edu); Dietram A. Scheufele, University of Wisconsin-

Madison (scheufele@wisc.edu); Bret R. Shaw, University of Wisconsin-Madison (brshaw@wisc.edu)

Ample research in public opinion has demonstrated that attitude formation can be highly contingent upon the cognitive structures that are most salient to an individual. Such priming mechanisms have been attributed to survey response (e.g., question-order) effects as well as issues that receive abundant attention in news media (e.g., agenda-setting effects). The present study moves beyond the context of everyday public-affairs news or political campaigns to investigate priming as it functions in attitude formation regarding emerging scientific and technological innovations. In particular, we focus on alternative energy in the form of biofuels, which have been proposed as a promising solution to the fossil fuel-based energy economy of the United States.

In order to investigate these mechanisms, we embedded an experimental manipulation in an RDD telephone survey of 600 residents of a top biofuels-producing state (conducted April to June 2009). Prior to assessing support for biofuels (both in terms of their impacts for society and the economy), respondents were primed to think of this issue in terms of consequences for (1) the environment, (2) the economy, (3) morals/ethics, and (4) political/governmental policy. Within each condition, primes consisted of evaluations of two potential risks and two possible benefits of biofuels. Our results revealed no differences in support for biofuels due simply to random assignment in one of the four priming conditions. However, we found a significant interaction effect between extremity of response to the risk prime (but, importantly, not to the benefit prime) and experimental condition on support for this technology. In other words, respondents' support for biofuels varied both within and between conditions depending upon how they responded to the domain-specific risk prime. Implications for attitude formation and the theory of priming are discussed.

An End to a Means: Partisanship, Policy Preferences and Global Warming.

Patrick Moynihan, Harvard University (pmoynihan@iq.harvard.edu); Gary Langer, ABC News (gary.langer@abc.com); Peyton M. Craighill, GQR (peyton_craighill@yahoo.com)

Recent polls have shown significant declines in the belief that global warming is occurring, particularly among conservatives and Republicans – but whether this represents a fundamental shift in the public's understanding of climate change, or some sort of political calculation, is an open empirical question. While previous public opinion research has explored the political and social determinants of (1) belief in whether global warming is occurring, and (2) policy preferences concerning global warming, the standard explanatory model has generally been limited to designating the former (belief global warming is occurring) as a predictor of the latter (policy preferences). What has received considerably less attention in the literature is the reverse causal logic – that is, whether or not views on government action to address climate change can predict belief in whether it's occurring. In this paper we examine the bi-directionality of this relationship, what other factors are most influential in shaping it and the nature of changes over time. We explore these issues using nationally representative ABC News and ABC News/Washington Post data as well other contemporaneous national datasets within the broader context of eroding Republican self-identification and increasing ideological polarization. The paper concludes with a discussion of the challenges of disentangling the logic and empirical evidence of causal order in cross-sectional data.

Fading Beliefs: An Examination of Diminished Belief in Global Warming Among Americans

Christopher Paul Borick, Muhlenberg College Institute of Public Opinion (cborick@muhlenberg.edu); Barry G. Rabe, University of Michigan (brabe@umich.edu)

During the last calendar year the number of Americans who indicate a belief in global warming appears to have dropped. Research by Pew, the Gallup Organization and ABC News and the Washington Post have all shown a smaller percentage of Americans expressing belief that temperatures on earth have been increasing. This decline in levels of belief in climate change has occurred after a decade in which residents of the United States demonstrated increasing acceptance of this phenomenon. While the consistency of results from national level surveys suggests that Americans are indeed less likely to believe in global warming than in recent years, the reasons behind the declining results is less clear. Numerous factors such as cooler summer temperatures, increased economic concern, and more intense partisan debates about policies to address global warming have been cited as reasons for the decline in belief in climate change; however there has been little empirical evidence to substantiate these claims.

In this paper results from 2008 and 2009 National Surveys on American Public Opinion on Climate Change and Policy Options are examined in order to shed light on the reasons for declining levels of belief in global warming. The findings from these surveys show a 7 point drop in the percentage of Americans who believe that the planet is warming, and diminished confidence in this occurrence among those that do believe in climate change. The results of the surveys also provide evidence about the relative impact of public trust in the media, scientists and personal experiences in contributing to the lower belief levels in the United States. The findings suggest that individual perceptions of cooler local temperatures and decreasing confidence in scientific evidence have contributed to the drop in belief levels among Americans, and in particular among those affiliating themselves with the Republican Party.

Public Opinion on Climate Change: The Role of Trust.

Sandy Marquart-Pyatt, Michigan State University (marqua41@msu.edu)

Climate change is one of the most pressing environmental issues of our times. Yet some uncertainty remains regarding how information sources shape public opinion on this issue. This paper examines influences on public opinion about climate change among the US general public using the General Social Survey. Using a path model, the sources of knowledge, worry and concern about climate change are examined. Particular attention is paid to the mediating role of trust in scientists as a factor influencing attitudes regarding climate change. Results show important influences of key socio-demographic factors, as anticipated from prior research. Results also reveal that trust mediates the relationship education and climate change concern. Implications of this research are discussed.

The Public's Views of Environmental Issues in an Economic Downturn

Leah Melani Christian, Pew Research Center for the People & the Press

(Ichristian@pewresearch.org); Eugene MacIntyre, ChrisMac Communications

(eugenemacintyre@gmail.com)

As the United States has experienced the worst economic downturn since the Great Depression, economic concerns have dominated the public's attention. How has the public's focus on the economy and jobs influenced perceptions about other policy issues, such as environmental protection? To address this question, we examine recent trends in public opinion about environmental issues.

Recent public opinion surveys report that environmental issues have slipped as a top policy priority for many Americans, and very few people cite environmental issues as one of the most important problems facing the country today. When asked specifically about prioritizing environmental protection over economic growth, far fewer now say that protecting the environment should be given priority than did so just a few years ago. In addition, a smaller proportion of Americans say they are willing to pay higher prices to protect the environment. But surveys also show that over the past few years the public's support

for environmental protection and concern about most environmental issues have remained high, and these views have changed little in recent years. Similarly, most Americans continue to say that we spend too little on improving and protecting the environment.

To provide a more long-term perspective, we also compare these recent trends in the public's views of environmental issues with those from the last thirty years. Public opinion has ebbed and flowed over this period as the salience of environmental issues, evaluations of environmental conditions and support for government action have changed. We focus on analyzing how current trends compare to the public's views during previous periods of economic downturn.

Survey & Measurement Challenges with Racial/Ethnic Populations

Survey Satisficing in Different Populations: To What Extent Do Response Effects Reflect Satisficing across Racial and Ethnic Groups?

Allyson L Holbrook, Survey Research Laboratory, University of Illinois at Chicago (allyson@uic.edu); Timothy P Johnson, Survey Research Laboratory, University of Illinois at Chicago (tjohnson@srl.uic.edu); Young Ik Cho, Survey Research Laboratory, University of Illinois at Chicago (youngc@srl.uic.edu); Sharon Shavitt, University of Illinois (shavitt@illinois.edu); Noel Chavez, University of Illinois at Chicago (nchavez@uic.edu); Saul Weiner, University of Illinois at Chicago (sweiner@uic.edu)

Survey satisficing occurs when respondents in a survey do not fully and carefully go through the cognitive processes necessary to optimize and provide the most accurate survey response (Krosnick 1991). Acquiescence response bias, no opinion responding, response order effects, and nondifferentiation have been proposed as reflecting common strategies respondents use to satisfice. Using data from a face-to-face survey, we explore whether each of these strategies reflects survey satisficing within each of three ethnic and racial groups: non-Hispanic Whites, African-Americans, and Mexican-Americans. To do so, we go beyond past research that has compared mean levels of these response strategies (e.g., Johnson et al. 1997) by examining whether within each group each strategy occurs more under conditions thought to foster satisficing. We do so by assessing within each group whether each strategy is more prevalent among respondents who are less motivated and able to optimize (e.g., among respondents low in cognitive skills, need for cognition, and late in the survey when respondents become fatigued).

References:

Krosnick, J. A. (1991). Response Strategies For Coping With The Cognitive Demands Of Attitude Measures In Surveys. Special issue: Cognition and survey measurement. *Applied Cognitive Psychology*, 5, 213-36.

Johnson, T. P., O'Rourke, D., Chávez, N., Sudman, S., Warnecke, R., Lacey, L., et al. (1997). Social Cognition And Responses To Survey Questions Among Culturally Diverse Populations. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality*. New York: John Wiley & Sons.

Motivations and Barriers to Participation in the MIDUS Biomarker Study: Comparisons Between a National Sample and a Minority Oversample.

Kristofer Hansen, University of Wisconsin-Madison, Institute on Aging (krhansen@ssc.wisc.edu); Nancy Davenport, University of Wisconsin-Madison, Institute on Aging (njdavenport@wisc.edu);

Gayle D Love, University of Wisconsin-Madison, Institute on Aging (glove@wisc.edu); Barry Radler, University of Wisconsin-Madison, Institute on Aging (bradler@wisc.edu)

Psychological, sociological, and health researchers have long been concerned about reaching and collecting data from the African-American population. In longitudinal studies, concerns are also growing about minimizing attrition. MIDUS, a longitudinal, mixed mode study of health and aging in a national sample, included a new city specific African American oversample, which provides unique opportunities to address these issues. This paper discusses factors influencing the decision to participate, or not, in one phase of MIDUS for both the National MIDUS sample and the African American oversample. Motivations and barriers to biological participation are examined for participants (n=935) and non-participants (n=1439), respectively. Significant differences in reasons for participating were found between the two groups. The National sample was more likely to report being motivated by a desire to be compliant or by the nature of the study, while the African-American sample was more likely to report being motivated by the opportunity to receive health reports or the monetary incentive. In contrast, a desire to learn was among the top three motivations reported by both groups. There are also significant differences in barriers to participation. The African American sample was more likely to cite barriers such as their health, while the National sample was more likely to identify travel as a barrier. The second most frequent barrier identified by both groups was family problems. There are also significant differences in the demographic characteristics (age, gender, education, income) of the two groups. Discussions will focus on how these results can be used for recruiting and retaining minority samples in longitudinal research, and how both reported motivations and barriers can be useful for maximizing participation.

Measuring Acculturation in the U.S. Hispanic Population

Wendy Mansfield, Knowledge Networks (wmansfield@knowledge networks.com); Jordon Peugh, Knowledge Networks (jpeugh@knowledge networks.com); Jannet Torres, Knowledge Networks (jtorres@knowledge networks.com); Tom Wells, Knowledge Networks (twells@knowledge networks.com)

With Hispanics now representing over 15% of the U.S. population, failure to sample Hispanics from the full range of the acculturation scale can have an impact on survey results. Using KnowledgePanel LatinoSM, an online web panel designed to be representative of the national Latino population, and using responses to 29 agree/disagree statements on lifestyle issues from a 2009 survey, this paper explores the accuracy of different acculturation measures and the effect of excluding the least acculturated group of Hispanics from samples.

How much improvement in the accuracy of acculturation measures is gained by moving from unidimensional measures of acculturation to a multidimensional measure? Our examination of three unidimensional measures (citizenship status, immigration generation, and years living in the U.S.) and one multidimensional measure (the Bidimensional Acculturation Scale for Hispanics [BAS]) quantifies the extent to which the BAS scale is superior to the other measures. Preliminary analysis indicates that the BAS scale is substantially more likely to define subgroups of Hispanics for which the more acculturated a subgroup is, the more similar its answers to lifestyle questions are to the answers of non-Hispanic U.S.-born citizens. The lifestyle questions cover work-life balance, leisure, socializing, retirement, the environment, and other topics.

How do survey responses to lifestyle questions differ when the least acculturated group of Hispanics is excluded from the sample? Preliminary analysis indicates that there is a statistically significant difference between the proportion of Hispanics and the proportion of non-Hispanic U.S.-born citizens agreeing with the opinion statements for 93% of the statements (across all 4 acculturation variables) when all Hispanic

groups are included, but for only 69% of the statements when the least acculturated Hispanic group is excluded.

The paper will present findings and discuss implications for using the different acculturation measures and for including all or only some groups of Hispanics.

Racial and Ethnic Group Response Rates: A Meta-Analysis of Techniques for Increasing Responses.

Lori Shelby, George Mason University (lshelby@gmu.edu)

Previous research is plentiful on methodological techniques for improving survey response rates. Methodological choices for improving survey response rates have included, for example, survey mode, privacy and confidentiality, incentives, topic salience, interviewer characteristics, survey appearance and design, postage type, and advanced notice and follow-up frequency. There is a growing body of research on the underrepresentation of racial and ethnic minority groups in survey results. Little attention, however, has been given to specific methods for increasing survey response rates for specific underrepresented racial and ethnic minority groups. This paper reports on a meta-analysis of prior methodological studies of techniques designed to increase survey response rates with a specific focus on studies that included race and ethnicity variables. Studies were chosen for inclusion based on a series of steps: (1) a comprehensive literature review of response rate research was conducted for all available years in digital databases (e.g., PsycINFO, Ingenta), (2) reference lists within the selected articles were checked for additional articles, and (3) a search of journals with a history of publishing survey response rate studies was conducted (e.g., Journal of Official Statistics, The Public Opinion Quarterly). Effect sizes for methodological choice predictors are examined in relation to the response rates for specific racial and ethnic minority groups. Other demographic variables were also recorded and analyzed (i.e., gender, age, education level, urban vs. rural, and socioeconomic status). Findings showing the key methodological choices for improving specific racial and ethnic group response rates are discussed. In addition, methodological areas (e.g., survey appearance) where applied research is exceptionally limited and necessary to help pinpoint best practices for maximizing specific ethnicity and racial group response rates is discussed.

Asian Group Threat and Malleable White Attitudes Toward Grade Point Average

Frank Samson, University of Miami (flsamson@miami.edu)

The principles of meritocracy and fairness sat at the center of changes to the University of California's admissions policies and a lawsuit filed by white firefighters against New Haven, Connecticut. I test white commitment to the values of meritocracy and fairness, in terms of the willingness to change the importance placed on grade point average for public university admissions in the context of ethnoracial out-group threat, using a laboratory experiment with a sample of 64 subjects at a highly selective university and a survey-based experiment of a sample of 440 California adult residents. The main finding reveals that whites decrease the importance of grade point average in response to experimentally manipulated prompts cueing Asian numerical dominance and university overrepresentation (relative to their proportion in the state). However, white Californians increase the importance of grade point average in response to perceived black group competition or when exposed to black and Asian group threats simultaneously. The results suggest that white support for meritocracy shifts depending upon the ethnoracial out-group threat. The results also suggest that in a multiracial context, an ethnoracial hierarchy exists in the minds of whites where blacks are deemed more threatening than Asians.

AAPOR Concurrent Session B

Friday, May 14, 2010

10:00 a.m. - 11:30 a.m.

Advances in Survey Sampling & Weighting

The Impact of Survey Design Modifications on Health Insurance Coverage Estimates in a National Longitudinal Health Care Survey

Steven B Cohen, Agency for Healthcare Research and Quality (scohen@ahrq.gov); Trena Ezzati-Rice, AHRQ (tezzatir@ahrq.gov)

National health insurance coverage estimates for the overall population and specific subgroups are critical to policymakers concerned with access to medical care and the cost and sources of payment for that care. The Medical Expenditure Panel Survey (MEPS) is one of the core health care surveys in the United States that serves as a primary source for these essential national health insurance coverage estimates. The survey was designed to provide annual national estimates of the health care use, medical expenditures, sources of payment and insurance coverage for the U.S. civilian non-institutionalized population. In 2007, the survey experienced two dominant survey design modifications: a new sample design attributable to the redesign of the National Health Interview Survey, and an upgrade to the CAPI platform for survey instrument, moving from a DOS to a Windows-based environment. This study examines the impact of these survey design modifications on the national estimates of insurance coverage.

The overlapping panel design of the MEPS survey and its longitudinal features are particularly well suited to assess the impact of survey redesign modifications on estimates. Since two independent nationally representative samples are pooled to produce calendar year estimates, one has the capacity to compare estimates based on the “original survey design” in contrast to those derived from the “survey redesign.” Attention is given to the correlates of nonresponse incorporated in the estimation techniques and adjustment methods employed in the survey and the measures utilized for post-stratification overall and by panel. Particular emphasis is given to assessing the level of convergence in coverage estimates based on the alternative designs as well as the alignment of model based analyses that discern the factors associated with insurance classifications. The presentation concludes with a discussion of strategies under consideration that may yield additional improvements in accuracy for these critical policy relevant survey estimates.

The American Community Survey (ACS) as a Sample Source For The National Immunization Survey (NIS): Results of an Evaluation Study.

Christopher Stringer, U.S. Census Bureau (M.Christopher.Stringer@census.gov); James Singleton, Centers for Disease Control and Prevention (JSingleton@cdc.gov); Andrea Piani, U.S. Census Bureau (Andrea.L.Piani@census.gov); Philip Smith, Centers for Disease Control and Prevention (PSmith3@cdc.gov); Karen Wooten, Centers for Disease Control and Prevention (KWooten@cdc.gov); Larry Wilkinson, Centers for Disease Control and Prevention (LHWilkinson@cdc.gov)

The Centers for Disease Control and Prevention and the U.S. Census Bureau partnered to conduct an evaluation study of the National Immunization Survey (NIS) using the American Community Survey (ACS)

as a sampling frame. One intended use of the ACS as a sampling frame is the identification of more difficult to reach populations, thus overcoming the substantial telephone screening required to identify households with children aged 19-35 months for the NIS and potentially reducing bias from non response in the survey. The primary sampling frame for this study was households who participated in the ACS and reported having children who would fall into the NIS target population at the time of the evaluation study. A dual frame was utilized on one geographic stratum, where a secondary, alternative sampling source was used to supplement the ACS cases in order to achieve the sample size requirements for the study.

The current NIS data collection consists of two phases. Sample households are contacted by telephone to identify those with age-eligible children and parents/guardians are asked to give permission to contact their vaccination providers. Vaccination providers are then mailed a survey to collect immunization data on sample children, in order to determine vaccination status. The household data collection plan for the evaluation study was expanded to a multi-mode approach that included in person interviewing. Field follow-up interviews were conducted for households who were unable to be contacted or refused to participate via telephone. Special provisions were required for the study to comply with confidentiality requirements mandated by Title 13 of the U.S. Code. This presentation explores the operational challenges associated with collection of provider vaccination data and the impact Title 13 may have on provider cooperation. Results of household interviewing are discussed along with preliminary results of the provider operation.

Reconstruction of Best Estimate Weights From Adjusted Microdata With Application to the 2004 NEP Exit Polls.

Jan Werner, Jan Werner Data Processing (jwerner@jwdp.com)

Major news media advocate transparency, but often conceal critical details of their own surveys. National Election Pool exit poll microdata are made available only as weighted to match the final election results, preventing any meaningful evaluation of the actual performance of those polls.

After accidental revelations that the 2004 exit polls incorrectly suggested a Kerry victory in Ohio, Edison/Mitofsky, who conducted the polls for NEP, issued a report (2005) describing errors, identifying operational problems, and promising corrective measures. Since then, NEP has been successful in enforcing its secrecy rules, making it difficult to determine whether there has been any improvement in exit poll performance.

This paper demonstrates how best estimate weights, prior to their adjustment to match the final election outcome, can be reconstructed using weighted and unweighted microdata. This was applied to the 2004 exit poll microdata released by NEP and the results are shown to agree substantively with details published in the 2005 E/M report. The method shows promise for evaluating the performance of other polls or surveys when weighted microdata are available.

Potentials and Constraints of Propensity Score Weighting to Improve Web Survey Quality

Stephanie Steinmetz, Erasmus University Rotterdam (steinmetz@fsw.eur.nl); Kea Tijdens, Erasmus University Rotterdam (tijdens@fsw.eur.nl)

With the introduction of the Internet as a new data collection mode, traditional survey methodology is challenged fundamentally. The increasing popularity of web surveys created a demand for appropriate web survey methodology, thereby triggering a heated debate about the quality and reliability of web surveys for scientific use. The most obvious potential drawback of web surveys is that they may not be

representative of the population of interest because the sub-population with Internet access is quite specific.

In this context, propensity score adjustment (PSA) has been proposed to statistically surmount inherent problems, particularly in non-probability-based web surveys. In this procedure, a parallel probability-based reference survey is used to estimate the propensities of being in the web sample based on a vector of covariates (socio-demographic and 'webographic' (lifestyle) variables) measured in both samples. It is critical for the method to choose appropriate covariates. In the scientific community, however, this method has not been applied traditionally in the field of surveys. There has been a minimal amount of evidence for its applicability. The implications particularly for survey methodology still need to be studied more extensively.

Against this background, the paper will explore the efficiency of PSA in adjusting biases arising from non-randomized sample selection. For that purpose, the un-weighted and weighted results from the Dutch volunteer web sample of the WageIndicator Survey 2009 will be compared with data from the Dutch LISS Panel that have been collected parallel. The advantage of this reference survey is that it provides a proper probability sample stemming from the same resource (e.g. the same questionnaire). Survey mode effects can be excluded as both questionnaires are completed individually on the computer. Finally, the application will also examine the sensitivity of the results, particularly with regard to changes in the specification of the propensity score.

Toward a Standardization of Survey Weights: The American National Election Studies Weighting System.

Josh Pasek, Stanford University (josh@joshpasek.com); Matthew DeBell, Stanford University (debell@stanford.edu); Jon A. Krosnick, Stanford University (krosnick@stanford.edu)

Although there has been a tremendous amount of publication on poststratification in recent years, there appears to be no universally agreed upon, straightforward method for survey weight construction. Indeed, many prominent practitioners believe that constructing weights should be an art, not a science. There is no agreed upon standard for determining which variables should be used, what types of algorithms should be employed, whether and how weights should be truncated, and a variety of other important considerations.

In order to identify best practices in constructing survey weights, the American National Election Studies (ANES) assembled a panel of experts in the field. Doug Rivers, Martin Frankel, Colm O'Muircheartaigh, Charles Franklin, and Andrew Gelman spent months working over a series of recommendations for how to best construct weights. Their work culminated in a September, 2009 report on how weights should be computed for the ANES and other similar studies.

In this presentation, we will describe the methods proposed in the ANES weighting report. We will also introduce a new software package designed to implement the agreed-upon procedures. The software package presents a simpler interface for specifying weighting information, comes pre-loaded with demographic information to match each year's demographics for the United States population from 2000 to 2009, and also automatically conducts analyses of which variables account for the largest discrepancies from known population parameters. In addition, the algorithm allows the user to specify mechanisms for variable selection, for the creation of weights on separate portions of the population, and for the creation of identically parameterized weights across multiple subsamples.

Challenges of Multimode Data Collection

Diversity of Methods: Assessment of Quantitative and Qualitative Research Multiplier Effects.
Janice Ballou, Mathematica Policy Research (jballou@mathematica-mpr.com); Brian Roff,
Mathematica Policy Research (broff@mathematica-mpr.com); Matthew Anderson, Mathematica
Policy Research (manderson@mathematica-mpr.com)

In his 2003 AAPOR Presidential Address, Mark Schulman challenged survey researchers "...to move beyond our desks and our libraries and make contact with the real world," and summarized his message noting "neither qualitative nor survey research has a monopoly on uncovering the hidden truths. They often have a multiplier effect when they are used together."

Research conducted for the Office of Research Integrity (ORI) provided the opportunity to both move beyond our desks and examine Schulman's multiplier hypothesis. ORI's mission is to prevent scientific misconduct and promote research integrity. To find out how faculty view their role in training future scientists to be responsible researchers, ORI sponsored a web survey of 3,534 recipients of 2005–2006 National Institutes of Health grants and a pilot test of in-person interviews with nine faculty members who had also completed the web questionnaire. We will use the data from these studies to compare the web survey responses with the in-person interviews to answer two research questions: (1) Are there response differences related to how faculty understands the questions and form their answers in the quantitative and qualitative experiences? and (2) What are the multiplier effects—how, if at all, is the quality of information enhanced when both qualitative and quantitative methods are used to study a topic? The topics we will compare related to educating doctoral students are: (1) faculty role descriptions, (2) graduate student education responsibilities, (3) activities conducted with doctoral students, and (4) support from academic departments and institutions for training students in the responsible conduct of research. To conduct this comparison we will use methodological research from conversational and cognitive interviewing and questionnaire development.

From Groundskeepers to Administrators: A Research Design for a Comprehensive Survey of a Diverse Employee Population.

Nancy Whelchel, N.C. State University (nancy_whelchel@ncsu.edu)

NC State University successfully developed a research design to gather information on a wide range of topics from its demographically and occupationally diverse employee population. The "2008 Staff Well-Being Survey" survey had a 60% response rate (AAPOR2) with little evidence of self-selection bias. Despite concerns about the extent to which employees in "blue collar" jobs would be interested and able to participate in the survey, the demographic make-up of respondents was nearly identical to the 5,800 staff in the population on measures including not only occupational classification, but also race/ethnicity, gender, age, and department/unit. For example, 6.1% of the population is classified as "skilled crafts" and 10.0% as "service/maintenance," compared to 6.3% and 9.5% of respondents. Hispanics and African Americans each make up 3.4% and 18.3% of the population, compared to 3.3% and 16.4% of respondents.

In this paper I detail strategies used to develop and administer the survey that could be helpful in addressing diversity issues in similar surveys. For example, the questionnaire was developed in consultation with an advisory committee including representatives from units across campus to insure that it covered issues important to those in a wide range of departments/occupations. It was pre-tested with groups from different job classifications to assure comprehensibility regardless of education level. The

survey was widely promoted through campus media, staff meetings, and various employee gathering places to increase awareness of the project. Participation among those with limited Internet access was facilitated by the addition of a paper survey to the online option. Similarly, correspondence was tailored so those in occupations classified as “skilled crafts” or “service/maintenance” were sent paper survey packets and postcard reminders through campus mail in addition to the standard e-mail communication. Various incentives (football tickets, theatre tickets, bookstore gift cards) were used to appeal to diversity in personal preferences.

Response Mode and Bias Analysis in the IRS' Individual Taxpayer Burden Survey

Karen Masken, Internal Revenue Service, Office of Research, Analysis & Statistics

(karen.c.masken@irs.gov); George Contos, Internal Revenue Service, Office of Research, Analysis & Statistics (george.contos@irs.gov); Roy Nord, Internal Revenue Service, Office Research, Analysis & Statistics (roy.d.nord@irs.gov); Michael Brick, Westat (mikebrick@westat.com)

The Internal Revenue Service (IRS) conducted a survey to measure the time and money that individuals spend on pre-filing and filing activities in response to the requirements of the U.S. federal tax system. The sampling frame was taxpayers who filed a 2007 income tax return in 2008.

The data collection protocol depended on whether the sampled taxpayer could be matched to a telephone number. Telephone numbers were found for approximately 76 percent of the sampled taxpayers and these were classified as ‘telephone matches’, the remainder are ‘nonmatches.’ Both groups were sent an initial mailing providing a detailed description of the purpose of the survey along with a letter from an IRS executive emphasizing the importance of the study and ensuring that the information collected would not be used for enforcement purposes. It also included a one-dollar bill as “an attention getter” and indicated that respondents would receive \$25 if they completed the survey.

In the initial mailing, the telephone matches were informed they could wait for a call from the survey administrator or complete the survey on-line by going to a specified URL. The “non-matches” group members were sent a letter that provided the web address (URL) and were told a mail questionnaire was being sent.

The overall response rate was 47.7 percent. Of the three-fourths that were telephone matches (76.2 percent), the response rate for the matched cases was 51.6 percent; the response rate for the nonmatches (23.8 percent) was 35.2 percent. One interesting finding is that a surprisingly high percentage responded by the web. Overall, 28 percent of all the responses were completed on-line, which is higher than in other data collection efforts that have been reported in the literature.

This paper explores characteristics of respondents and non-respondents and highlights those who responded on-line.

To Mail or to Call: How to Reach the Hard-to-Reach.

Robin Gentry, Arbitron (robin.gentry@arbitron.com); Michelle Cantave, Arbitron, Inc.

(michelle.cantave@arbitron.com); Nicole Wasikowski, Arbitron, Inc.

(nicole.wasikowski@arbitron.com); Yelena Pens, Arbitron, Inc. (yelena.pens@arbitron.com)

Commercial databases are widely used to match addresses to sampled phone numbers or phone numbers to sampled addresses. This approach allows the researcher to mail pre-alerts to RDD sample or complete follow-up calls with address based sample—hopefully improving contact and cooperation rates.

However, due to database limitations, often both an address and phone number are not available for all sample records. Arbitron, a provider of media ratings, uses a hybrid sampling frame—a combination of RDD phone and address-based sample to recruit households for its syndicated radio listening survey. This mixed mode approach allows an opportunity to determine the most appropriate initial method (phone or mail) for contacting sample which does not have both phone and address information. In theory, unmatched landline households recruited via either approach—address or phone—should be from the same population because the same commercial databases were used to do the matching. With the RDD approach, households without an address match are called and asked to participate in a one-week radio ratings diary survey. Because there is no available address, no pre-alerts and/or incentives can be sent prior to calling. In contrast, addresses without a phone number match are mailed a questionnaire and small incentive and are asked about media habits, household demographics and phone contact information. They are also offered a promised incentive for returning the questionnaire. Following this screener survey, households are re-contacted by telephone to gain their participation in the same one week radio diary survey. This presentation will compare the demographic characteristics and radio listening habits of respondents recruited from both methods and will offer recommendations about the optimal approach to use when trying to reach households with limited contact information at the time of sampling.

Impact of Monetary Incentives and Web Survey Option in the 2008 National Survey of Recent College Graduates (NSRCG) on Increasing Response Rate in Historically Low Responding Groups.

Sheila Heaviside, Mathematica Policy Research (sheaviside@mathematica-mpr.com); Donsig Jang, Mathematica Policy Research (djang@mathematica-mpr.com); Geraldine Mooney, Mathematica Policy Research (gmooney@mathematica-mpr.com); Kirsten Barrett, Mathematica Policy Research (kbarrett@mathematica-mpr.com); Kelly Kang, National Science Foundation (kkang@nsf.gov)

Historically, the National Survey of Recent College Graduates (NSRCG) has noted differences in survey response rate by sample members' race/ethnicity, U.S. residency status and academic major. In 2006, for example, the unweighted response rates for minority sample members averaged about 62 percent, compared to 75 percent for non-Hispanic white sample members. Also in that same year, only 51 percent of sample members with foreign addresses completed the survey, compared to 70 percent of those with U.S. addresses. Differences of five or more percentage points were also found across academic majors.

The NSRCG is a multimode survey of approximately 18,000 randomly selected recent college graduates. In 2008 NSRCG, a web survey option was made available for the entire duration of the survey for the first time. Also in 2008, NSRCG embedded an experiment that offered variations in timing and amount of incentives and a web response option. Sample members were randomly assigned to various incentive and web option groups. This paper examines impact of incentives and web response option on increasing the survey completion rates among the sample groups with historically low survey responses. It takes advantage of an incentive experiment built into the study and compares rate of response among these subgroups by incentives offered and use of the web option.

Issues in Cell Phone Surveys

Experiments in Cell Phone Nonresponse

Andy Peytchev, RTI International (andrey@umich.edu); Karol Krotki, RTI International (kkrotki@rti.org)

Inarguably, interactions over cell phones are different from those over landlines. There are multiple factors that can uniquely impact survey requests over cell phones, including (1) the expected absence of unsolicited calls and (2) the financial responsibility of incoming calls. Since only the phone number is available in cell phone random-digit-dialing studies, limiting the channels of communication with sample members (e.g., no prenotification letters and no alternate landline numbers), practitioners have little control over these factors.

We carried out two experiments to reduce the impact of these two factors on nonresponse. Providing study information in voicemail messages can supply sample members with information about the reason for the call and allow them to expect the call – potentially decreasing nonresponse. It may, however, provide the opportunity to make a nonparticipation decision prior to talking with the interviewer, thereby increasing nonresponse. A random half of the sample was assigned to receive voicemail messages on the first call attempt if contact was not achieved.

Yet even when contacted, sample members may refuse in part due to costs incurred for “airtime.” To alleviate this concern, sample members could be given the option to be called on another number. Any reduction in nonresponse, however, may be lost and nonresponse even increased if potential respondents are difficult to contact or refuse when called on the alternate number. In this second experiment, sample members were randomly assigned to receive an option to be called on another number.

The study was fielded in June and July, 2009, on national samples of cell phone numbers from MSG and from SSI. Preliminary findings show weak indication of favorable effect of leaving voicemail messages with the reason for the call and a strong indication of adverse effect of providing an option to be called on another number.

Small-Area Cell Phone Only Estimates

Anna Fleeman, Arbitron Inc. (anna.fleeman@arbitron.com); Richard Griffiths, Arbitron Inc. (richard.griffiths@arbitron.com); Robin Gentry, Arbitron Inc. (robin.gentry@arbitron.com); Nicole Wasikowski, Arbitron Inc (nicole.wasikowski@arbitron.com); Amita Dahra, Arbitron Inc. (amita.dahra@arbitron.com); Richard Possett, Arbitron Inc (richard.possett@arbitron.com)

As of the Fall 2009 Radio Ratings Survey, Arbitron has sampled Cell-Phone-Only (CPO) households in all markets (~300) across the U.S. These CPO households are identified using an address-based sample frame. Once the addresses are randomly selected, phone numbers, if available, are matched to them. For the addresses to which a phone number cannot be matched, a paper questionnaire is mailed. The questionnaire asks about media-related behaviors, cell/landline ownership, demographics and requests contact information. About one-third of the questionnaires are returned, and of these ~40% are CPO households. The zip codes of these CPO households are then geo-coded to a county and assigned to an Arbitron-defined market. To reduce possible non-response bias and attenuate any sample size issues, the NHIS state-level CPO penetration rates are used in conjunction with Arbitron’s results to produce the small-area CPO estimates. Small-area CPO estimates are of paramount importance to survey research as the number of CPO households continues to increase and the need grows for CPO estimates on local or regional levels for sampling and/or weighting purposes. Over the past 15 months, Arbitron has mailed more than one million questionnaires to selected addresses across the country and more than 100,000 CPO households have been identified via this methodology. Findings to be presented include noteworthy differences among Census regions and large metropolitan areas, intuitive areas of high CPO penetration,

such as in “college towns” or near military installations, and whether the national-based correlates of CPO status – age, gender, ethnicity, and household size – apply to smaller geographies.

The Variable Costs of Cell Phone Interviewing: Understanding Cost and Productivity Ratios in Dual-Frame Telephone Surveys.

Thomas M. Guterbock, Center for Survey Research, University of Virginia (tmg1p@virginia.edu); Paul J Lavrakas, Stamford, CT (pjlavrak@optonline.net); Trevor N. Thompson, Associated Press (TTompson@ap.org); Randall ZuWallack, Macro International (randal.zuwallack@macrointernational.com)

It is well known that the per-interview cost of telephone interviews conducted by means of randomly dialed cell phone numbers is higher than the cost of landline RDD interviews. However, the ratio of cell phone to landline interviewing costs varies widely across studies, for reasons that have been difficult for practitioners to predict.

Through our work as a subcommittee of the 2009 AAPOR Cell Phone Task Force, we have gathered detailed cost and productivity information on several dozen dual-frame telephone studies conducted recently by academic and commercial organizations. These data allow us to estimate average cost ratios for dual-frame studies that screen for cell-phone-onlies and for those that interview all cell phone users reached. Within each type of study, the cost ratios vary widely, and we are able to relate these variations to differences in hourly productivity and to other costs.

Relative hourly productivity depends on whether or not dual-phone users are screened out, the specificity of sample geography, variations in working number density on the cell phone and landline sides, variations in interview length, the specificity of screening criteria used in the study, differences in incidence rates, and the efficiency of different dialing technologies. Other cost factors include differential incentive costs and sample costs. We consider modeling strategies that would allow practitioners to predict with greater insight the cost of cell phone calling in future studies.

Assessing Coverage Error in an RDD Survey

Kathleen B. Santos, NORC at the University of Chicago (santos-kathleen@norc.org); Kennon R. Copeland, NORC at the University of Chicago (copeland-kennon@norc.org); Philip J. Smith, Centers for Disease Control and Prevention (pzs6@cdc.gov); Kirk M. Wolter, NORC at the University of Chicago (wolter-kirk@norc.org)

The issue of undercoverage error in random-digit-dial (RDD) frames has been growing in importance due to the rapid increase in cell phone usage. Blumberg, et al (2009) have reported the proportion of households with access to only wireless telephones (“wireless-only” households) has risen from roughly 7% in early 2005 to just over 20% in late 2008.

The National Immunization Survey (NIS) – a nationwide, list-assisted RDD survey sponsored by the Centers for Disease Control and Prevention and fielded by NORC – monitors the vaccination rates of children between the ages of 19 and 35 months and adolescents between the ages of 13 and 17 years. Thus, the NIS sample design excludes wireless-only households as well as households with no telephones. While ratio adjustment is implemented in the survey weighting to control for population undercoverage, there is the potential for coverage bias to be present in the NIS estimates.

The NIS-Registry project was designed to examine potential coverage, response, and provider nonresponse error in the NIS. A sample of NIS-eligible children was selected from the state immunization

registries in Arizona and Michigan. The NIS household and provider-record-check (PRC) questionnaires were administered for both samples during Q1/2008 and Q2/2008. The goal of this paper is to assess the nature and impact of differential coverage between the two sample frames – the NIS RDD frame, which covered only children in households with a landline telephone, and the state registry frame, which conceptually covered all children. The analysis will compare sociodemographic characteristics reported during the household interview as well as vaccination rates as reported by the provider through the PRC for the NIS and NIS-Registry samples to inform about coverage error in the NIS.

Memory, Recall & Accuracy in Survey Measurement

The Extent of Parallel Memories in Autobiographical Knowledge and its Manifestation in Calendar and Conventional Surveys

Robert F Belli, University of Nebraska (rbelli2@unl.edu); Tarek Albaghal, University of Nebraska (mtalbaghal@hotmail.com); Clarissa Steele, University of Nebraska (ivoace@yahoo.com)

Survey research is frequently interested in autobiographical information regarding varied events in an individual's life. Although current understanding of autobiographical knowledge shows that these memories are retrieved using top-down, sequential, and parallel processes, conventional questionnaire (CQ) surveys most often encourage top-down processes and a limited number of sequential ones. Event history calendar (EHC) surveys, however, are structured to cue all of these cognitive processes. Current models concerning the structure of autobiographical knowledge lack clarity on the frequency of parallel associations and the extent to which these occur in EHC and CQ surveys. Audio-taped EHC and CQ survey interviews were coded based on verbatim response to identify recalled parallel events across different autobiographical domains. Parallel memories were further coded for both theme and memory type. Analyses show that spontaneous parallel associations are a relatively common occurrence in surveys. More importantly, these memories are recalled more frequently in EHC compared to CQ surveys. Parallel memories were rarely specific episodes; rather, memories were either of a general or extended-event type. Log-linear models were tested to estimate the relationships between mode, autobiographical domain of the survey question, parallel theme, and memory type. Preliminary analysis shows that respondents in both survey types recall parallel associations between educational, labor, and residential events and their relationship histories, and remember educational, relationship, and residential events with their labor histories. General or extended-event parallel memories occur at approximately similar rates across domains, although extended-event memories occur somewhat more frequently for residence-themed parallel memories. These relationships are mediated by the three-way associations between domain, theme, and memory type. These analyses support the notion that models for autobiographical memory need to deemphasize the existence of hierarchical structures to account for the common presence of parallel associations. Further, researchers should incorporate new understandings of memory while developing questionnaires eliciting autobiographical knowledge.

Recall Error in Retrospective Survey Data.

Gerrit Mueller, IAB - Institute for Employment Research (gerrit.mueller@iab.de)

Surveys often rely on respondents' reports based on their autobiographical memory. Therefore, survey data may suffer from measurement error due to recall errors. Ignoring such error can affect parameter estimation and model-fit of models based on these data. The German panel survey PASS collects information on work histories using repeated interviews and retrospective reporting. PASS data have been linked to administrative data on work histories available from the German Federal Employment

Agency. As administrative data for these indicators have high validity, the paper focuses on recall error with respect to receipt of unemployment benefit and unemployment assistance payments. The first section contains a description of distributions of different kinds of errors found in the data. A second section deals with “seam effects” which are a very common type of error in survey research. A seam effect occurs when status changes are concentrated on the seam between the reporting periods of two consecutive waves of a survey. This can happen due to deliberate or accidental failure to report an event in any of the two waves or due to forward or backward telescoping (i.e. the tendency of people to remember events as occurring more or less recently than they actually did). The combined PASS and administrative data allow distinguishing between the two mechanisms. As PASS has a long field period of almost eight months and thus time since the last interview can vary, we can investigate the effect on recall error of time elapsed since the previous interview. In the last section we turn to the most important question if the likelihood of a recall error occur can be predicted by individual attributes like age or education and, if so, which effect recall error has on the coefficients of time-to-event models for the dynamics of unemployment assistance receipt.

The Use of Landmark Events in EHC-Interviews to Enhance Recall Accuracy

Wander van der Vaart, Wageningen University (w.vandervaart@chello.nl); Tina J. Glasner, Utrecht University (T.Glasner@uu.nl); Robert F. Belli, University of Nebraska (rbelli2@unl.edu)

In recent years there is an increasing use of calendar methods, like timeline procedures and Event History Calendars that aim to stimulate recall during survey interviews. Though research findings demonstrate that calendar instruments may improve recall accuracy, the results are mixed and it largely remains unclear how specific design characteristics of the calendar instruments add to the effects.

This paper aims to contribute to the further development of calendar instruments by examining one of its central characteristics: the use of landmark events. These are very salient events from people's lives that aim to facilitate recall of other events and particularly the dates of these events. Theoretical notions in this paper suggest that landmarks are most effective as recall aids if they are important, domain related, personal events.

The data in this study originate from the Event History Calendar as embedded in the PSID-survey in 1998. Using the recorded, transcribed and coded landmark events, this paper examines what types of personal landmarks were generated by the respondents and whether the use of these landmarks is related to recall accuracy. Data from the same respondents collected one year earlier were employed as a standard of comparison in order to assess the quality of the retrospective reports.

The findings demonstrate that respondents used a great variety of landmark events, that the number and types of landmarks were related to socio-demographic factors and that the retrieval of landmarks showed typical recency and heaping patterns. In the present study no clear relationships between landmark usage and recall accuracy was found. The discussion on the results suggests that a standardization of the landmark procedure in an EHC might add to the effectiveness of its aided recall function.

Unmotivated Anonymity: Social Desirability, Accuracy, and Satisficing under Conditions of Anonymity.

Yphtach Lelkes, Stanford University (ylelkes@stanford.edu); Jon Krosnick, Stanford University (krosnick@stanford.edu); Charles Judd, University of Colorado--Boulder (Charles.Judd@colorado.edu); David Marx, San Diego State University (dmarx@sciences.sdsu.edu)

Researchers often assume that promising anonymity to survey respondents shields results from social desirability bias. Anonymous surveys yield higher levels of socially inappropriate responses, but no studies have shown that responses gathered under anonymous conditions are also more accurate. Instead anonymity may act as a double-edged sword. Although anonymity decreases a person's motivation to engage in self-presentational motives, it also decreases accountability, thereby decreasing motivation to thoughtfully read and respond to questions. Two studies reported here demonstrate that anonymous participants provide less accurate responses than identifiable respondents. In the first study, participants researched a topic on the web, and then, after an anonymity or identifiability manipulation was delivered, were surveyed on their experience. The design of this experiment allowed us to assess actual versus reported web surfing behavior, time spent on the survey, and non-differentiation of response options between anonymous and identifiable conditions. In a second experiment, participants in the identifiable and anonymous conditions were questioned on various behaviors they may or may not engage in, including drinking, exercising, and sexual behaviors. The design of the questionnaire allowed us to determine if anonymous participants engaged in response behavior that would indicate that they were devoting fewer cognitive resources than identifiable participants, such as higher levels of acquiescence, non-differentiation, and selecting "don't know" options or writing less on open-ended questions. Although the studies replicate the social desirability findings of past research, anonymous participants also engaged in behavior that indicated that they were not as engaged in reading and responding to questions as identifiable participants. In addition, the experiments indicate a larger discrepancy between true behavior and reported behavior in anonymous conditions than in identifiable conditions. These studies suggest that anonymity may not be a costless panacea to the social desirability problem.

The effect of 'Do-you-remember' Probes in Event History Calendar Interviews

Yfke Ongena, University of Twente (y.p.ongena@utwente.nl); Wander Van der Vaart, Wageningen University (wander.vandervaat@wur.nl); Robert F. Belli, University of Nebraska (bbelli2@unlnotes.unl.edu)

Event History Calendar (EHC) interviews have been developed to improve respondent's recall of retrospective information, and are becoming a popular replacement of conventional (standardized) interviews. The current study contributes to further development of calendar interviewing approaches by systematically examining sequential patterns of interviewer-respondent verbal interactions in EHC interviews, and their effects on data quality.

Contrary to conventional interviews, EHC interviews give interviewers freedom to probe in their own words. Conversational interviewing styles in EHC interviews may also provoke narrative styles that may stimulate recall of autobiographical information. However, the extent to which interviewers should be allowed such freedom is subject of an ongoing debate on standardized versus conversational interviewing. This debate was primarily focused on whether interviewers should be allowed to clarify questions, but not on what specific memory cues interviewers can provide in their probing. Since little is known about the way interviewers utilize their freedom to probe and how this may influence respondent behaviour, this study focuses on spontaneous probes and the resulting quality of the answers obtained. Sequential analysis of interviewer-respondent interaction was performed on 120 transcripts of EHC interviews of the Panel Study of Income Dynamics.

One main finding is the differential use of timing probes. Instead of the direct question "and when did X happen?" about 15 percent of the timing probes are spontaneously formulated as "do you remember when X happened?" However, probes asking for specific elements (e.g., "What was the street name?")

are less often converted into 'Do-you-remember' probes (i.e., "Do you remember the street name?"). Analysis showed that these probes were related to data quality: respondents are more than three times as likely to give a don't know answer after 'Do-you-remember' probes than after direct probes. More findings demonstrate that interviewers probing styles are related to the quality of the interviewer-responder interaction.

Methodological Briefs: Improving Multimode Data Collection

Probability of Selecting Response Mode Between Paper- and Web-Based Options: Logistic Model Based on Institutional Demographics.

Marina Murray, Graduate Management Admission Council (mmurray@gmac.com)

During the last decade, a rapid growth in online surveys supported by advances in survey software technology and increased Internet usage inspired a great deal of methodological research to gain a better understanding of this survey mode. Most of this methodological research, however, focuses on individual respondents who participate in surveys to the general public. Are online surveys ready to replace paper-and-pencil formats for the purpose of studying organizations? When organizations have a choice, which method do they prefer? What factors influence a company's decision in the survey mode selection? The findings addressing these questions are based on primary research data collected by the Graduate Management Admission Council through its 2006 Diversity Survey of business schools in the United States. In addition to the traditional paper-and-pencil survey format, these US business schools were offered the alternative of completing the survey online. Using logistic regression, this research aims to estimate which institutional demographic characteristics influence the survey mode choice in establishment surveys. Depending on institutional demographics, which are typically known to the researcher, the results may be helpful in choosing which survey mode to utilize when conducting survey research of organizations.

Long-Term Efficacy of Sequential Mixed-Mode Designs on Response Rates and Cost in a Panel Survey.

Rachel M. Levenstein, University of Michigan (mlev@isr.umich.edu); Jennifer Barber, University of Michigan (jebarber@isr.umich.edu); Heather Gatny, University of Michigan (HGatny@isr.umich.edu)

Declining response rates in both cross-sectional surveys as well as panel surveys are of great concern to survey researchers (Groves & Peytcheva, 2008). One increasingly common method of increasing response rates is to switch the mode of administration in order to recruit nonrespondents (de Leeuw, 2005). Most often, these sequential mixed-mode designs use a relatively inexpensive mode for the initial contact attempts, then switch the mode at follow-up to a more expensive mode. While the use of sequential designs and their impacts on response rates are well-documented in cross-sectional surveys, we know little about their utility with respect to response rates and cost in panel surveys. The Relationship Dynamics and Social Life (RDSL) study is a panel survey that uses a design that initiates each wave in web and follows up nonrespondents in CATI at each wave. Web response rates tend to be lower than CATI response rates (Dillman et al., 2008), although the costs of CATI tend to be far greater than web (Couper, 2005). Thus, RDSL aims to balance cost and response rate concerns by only using CATI for the more difficult respondents. But panel respondents may be exposed to this sequential design, altering future behavior. This study will examine the impact of RDSL's sequential design on longitudinal response rates across modes, changes in the mode of completion, and changes in cost.

Mode Effect Analysis: Paper respondents vs. Web respondents in the 2004-05 Teacher Follow-up Survey.

Paola Tami Aritomi, American Institutes of Research (paritomi@air.org); Jason Hill, American Institutes of Research (jhill@air.org)

In order to address concerns that changes in future collection modes could impact the consistency of Teacher Followup Survey (TFS) estimates over time, the authors conducted a mode effect analysis on the 2004-05 administration. Although predominately a paper collection, the 2004-05 sampling design incorporated a small web-collection component. Using the experiment data for secondary analysis, the authors tested the 2004-05 TFS estimates for the possibility of mode effects. The more recent 2008-09 TFS data collection used paper surveys mainly to convert nonrespondents and for a small group of teacher who did not provide an e-mail address and were not sent the web survey. This study aims at exploring potential differences in teachers' responses between those who used the web questionnaire and those who opted for the paper questionnaire. The authors tested mode effects using six survey questions with different characteristics and levels of complexity. Selected questions with multiple items and multiple rating categories were transformed to be analyzed as estimated measures of differentiation. A two-stage Heckman-type instrumental variable (IV) regression model was used for these analyses. The first stage models whether teachers with certain characteristics were more prone to choose the web instrument than the paper instrument. The second stage of the model determines whether having used the web or paper instrument affected the quality of the survey responses, using the IV web-choice estimated from stage one. Regression results indicate that using the web-based instrument does not lead to lower quality or different survey responses compared to paper-based responses. As a result of the findings that support the initial hypothesis that no mode effects are observed on the 2004-05 TFS, the authors conclude that changes to data collection methodologies in the recently collected web-based 2008-09 TFS are unlikely to create long-term inconsistency issues.

The Effectiveness of Multiple Modes Among Young Adults

Mindy Anderson-Knott, University of Nebraska-Lincoln (mandersonknott2@unl.edu); Stacia Jorgensen, University of Nebraska-Lincoln (shalada2@unl.edu); Jeff Armitage, Nebraska Department of Health and Human Services (jeff.armitage@nebraska.gov)

A significant challenge for current survey research is gaining participation among young adults. The aim of this study is to increase participation of young adults by implementing a mixed mode (mail/web) approach. While some research on offering mixed modes has suggested a decrease in response rates (Dillman et al., 2008; Gentry, 2008; Griffen, Fischer, & Morgan, 2001; Grigorian & Hoffer, 2008), McCabe et al. (2006) suggests that utilizing a mail/web mixed mode approach could be effective at increasing response rates among some sub-populations, such as male college students. The timing of offering an alternative mode will be tested in this study using a survey measuring attitudes and behaviors regarding alcohol use. The survey will be administered using a 3 stage mailing to a regionally stratified statewide sample of 19-25 year olds in Nebraska. The sample will be randomly divided into four groups: 1) mail option only for all 3 mailings, 2) mail option only at 1st mailing, and mail with web option offered in the 2nd and 3rd mailings, 3) mail option only at 1st and 2nd mailing, and mail with web option offered in the 3rd mailing, and 4) mail with web option offered in all 3 mailings. Response rates will be compared to determine the effects of offering an alternative mode at different stages.

Emulating Interviewers in an Online Survey: Experimental Manipulation of 'Do-Not-Know' Over The Phone and on the Web.

Joop Hox, Department Methods & Statistics, Utrecht University (j.hox@uu.nl)

Web surveys, being self-administered, ensure more privacy for the respondent and there is considerable evidence that web surveys result in less socially desirable answers. However, the absence of an interviewer may also be seen as a disadvantage, as there are no interviewers to motivate respondents and probe them.

In this paper we report on an experiment in which a particular type of question format, i.e. how the “do not know” option is presented and whether or not a probe is used, affects answers to web and telephone surveys.

In interview surveys, usually a ‘do-not-know’-option is not explicitly offered to a respondent, but interviewers can accept it. It is considered good practice to train interviewers in using a probe after an initial ‘do-not-know’ to reduce item-nonresponse.

In web surveys designers are hesitant to offer an explicit do-not-know option for fear of encouraging respondents to choose this option as a quick answer. On the other hand, not accepting do-not-know and issuing an error message insisting on an answer, may lead to either irritation and more break-offs or to guessing and less valid answers.

Using the interactivity of the web, we emulated friendly interviewer probing behaviour in a probability based Internet panel. The questionnaire contained a series of questions, which in previous self-administered (mail and web) surveys showed a high percentage of item-nonresponse. A two by two experimental design was used: (1) explicit offering of do-not-know vs. no do-not-know option, and (2) directly accepting a do-not-know vs. only accepting it after a friendly probe. As baseline for comparison a fifth condition was added with the ‘standard’ web option: an error message with no acceptance of continuation without an answer. The number of resulting do-not-know answers in each condition are evaluated and compared with the results of a telephone survey on the same topic.

Visual vs. Aural Surveys: Examining the Similarities and Differences Between Survey Modes.
Carey Stapleton, Service Management Group (cstapleton@servicemanagement.com)

Using customer satisfaction survey data collected from over 400,000 respondents across 7 different multi-unit restaurant and retail brands, a complete examination of response differences between visual and aural surveys will be discussed. Respondents for each of the 7 brands were given the opportunity to respond to a 5 – 8 minute long survey accessing either a self-administered Interactive Voice Recording (IVR) or a website. With this research design and respondent sample, we will discuss differences in demographic profiles, break-off rates, sensitivity to questionnaire length, survey response scores, general item and sensitive question non-response, and response differentiation by survey mode.

Surveying on Sensitive Topics

Differences Between Confidentially and Orally Administered Overt Racism Measures: Evidence From the 2008 ANES.

Gaurav Sood, Stanford (gsood@stanford.edu); Jon Krosnick, Stanford University (krosnick@stanford.edu); Matthew DeBell, Stanford University (debell@stanford.edu)

Since 1992, every presidential year, in the post-election survey, the American National Election Studies have orally administered questions measuring people's perceptions of how hardworking and intelligent they think blacks and whites are. These questions seem to measure racial prejudice and to be politically consequential - they predict support for segregationist policies such as miscegenation laws,

opposition to general government assistance to blacks, and increase in opposition to popular opportunity-enhancing programs such as enterprise zones. However, some researchers have expressed concerns about the validity of such measures, because of concern that social desirability pressures may inhibit many Americans from honestly expressing their racist attitudes aloud to an interviewer during a face-to-face survey. To assess the validity of these criticisms, during the 2008 ANES, the racial stereotypes questions were asked using Computer Assisted Self Interview (ACASI- the method that has been extensively used to ask potentially sensitive questions in other surveys) during one interview and orally during another interview. Using data from this experiment and from other surveys that asked the same questions orally at about the same time of representative national samples, we explored whether confidential reporting using ACASI yielded more expression of anti-black prejudice by white respondents and more validity of the reports of perceived group qualities. The results suggest that while in-person interview does significantly downwardly bias the overall estimate of racism, it doesn't yield any less validity.

Racial Attitudes, the List Experiment and Hard to Reach Respondents: Measuring Reluctant Participants' Prejudice.

Michael Cobb, NC State University (mike_cobb@ncsu.edu)

This study examines the interaction between two types of survey error. First, it is widely believed that a non-trivial percentage of white respondents misreport their racial attitudes in surveys. Second, samples might not be representative because response rates have declined significantly over time. In this paper, I examine how the probability of respondent cooperation is related to the valid measurement of racial attitudes. I use data from a Pew experiment in 2003 that varied how rigorously respondents were recruited and that also included both direct and less obtrusive measures of racial attitudes. Preliminary findings support the contention that that direct measures of prejudice underestimate its prevalence, but also and ironically that the absolute level of prejudice is higher among samples obtained through standard methods than samples of respondents recruited through much more rigorous methods. I discuss the implications of these findings for past and future surveys about prejudice and other sensitive topics.

Breakdowns in the Cognitive Response Process as Potential Mechanisms Behind Measurement Error in Reports of Intimate Partner Violence

HarmoniJoie Noel, University of Nebraska-Lincoln (hnoel3@gmail.com)

Using a sample of 1,300 heterosexual couples from the National Longitudinal Study of Adolescent Health (Add Health), this paper explores the causes of measurement error in reports of intimate partner violence among young adults. Measurement error occurs when one partner reports physical violence in their relationship but the other partner does not. My study uniquely applies the cognitive response process used in the survey methodological literature to the partner violence literature to understand why partners disagree about violence in their relationship.

Differences in how men and women comprehend these questions, and retrieve and edit their answers may help explain disagreement about partner violence. To do this, proxies for each step of the process are identified in the Add Health data. For example, relationship duration may be a proxy for errors at the retrieval stage because couples who have been together for a shorter amount of time might be more accurate in their recall of violence than couples who have been together longer. I look at two types of physical violence – male perpetrated violence in which 12% of the couples disagree about violence and female perpetrated violence in which 22% of the couples disagree. For both types of violence, multinomial logistic regression models are used to examine male and female under and over reporting. By identifying and understanding the causes of disagreement, survey methodologists and partner violence researchers

can work to reduce or account for disagreement in order to improve the accuracy and reliability of partner violence estimates.

Researching Maternal Mortality Rates in Afghanistan

William C Hayes, D3 Systems, Inc. (will.hayes@d3systems.com); Niamh Darcy, RTI International (ndarcy@rti.org)

This paper focuses on processes used and unique challenges in implementing data collection in a maternity hospital in Afghanistan. It addresses the unique challenges presented by working on the private and sensitive topic of childbirth and accurately measuring training program results and post training implementation in an environment where armed conflict and gender-related political violence is common.

The United States Department of Health and Human Services (DHHS) contracted RTI International to conduct an evaluation of two training programs of the Afghanistan Health Initiative (AHI) implemented at the Rabia Balkhi Maternity Hospital (RBH) in Kabul. The AHI purpose at RBH was to improve the quality of maternal and newborn care through in-service training and systems strengthening.

Afghanistan has one of the highest maternal mortality rates in the world. Preventable complications related to childbirth cause more than 85 percent of deaths among women of childbearing age in Afghanistan. D3 research indicates that 25% of Afghan women have their first child before the age of 18, and one in three women believe they have been denied healthcare due to their gender.

The evaluation focused on clinical and hospital management training at RBH. RTI International led the evaluation design and partnered with D3 Systems to conduct data collection activities at the RBH facility. Several instruments were designed and adapted by RTI including an evaluation to assess the knowledge, skills and attitudes of the medical providers at RBH. RTI and D3 Systems' medically trained Afghan staff conducted clinical observations, qualitative interviews, a facility audit and observation of clinical trainers. These staff, who were mostly female, provided valuable insight into the practicalities of completing research in this unique context.

Focus Groups, Cognitive Interviews and Card Sorting: Three Methodologies to Improve Question Development for Different Cultural and Linguistic Groups

Susan Sprachman, Mathematica Policy Research (ssprachman@mathematica-mpr.com); Sally Atkins-Burnett, Mathematica Policy Research (satkins-burnett@mathematica-mpr.com); Nikki Aikens, Mathematic Policy Research (naikens@mathematica-mpr.com); Margaret Caspe, Children's Aid Society (margaretc@childrensaidsociety.org)

In the Universal Preschool Child Outcomes study, we used parents' rating of their children's behavior to assess children's socio-emotional development. Although these ratings are commonly used, we know that parents from different cultural and language backgrounds vary in their willingness to discuss particular behaviors. This paper explains our methodology for determining which standard scales were most appropriate for the development for the diverse ethnic groups in the study.

Our methods integrated three strategies – focus groups, card sorts, and cognitive interviewing – to inform the reliable and valid assessment of children's socio-emotional behavior via parent report.

We conducted six focus groups with 57 parents from diverse backgrounds (i.e., Filipino, Korean, Chinese, African American, bilingual Hispanics and monolingual Hispanics), the key groups in our study. Parents were asked to discuss the social skills and problem behaviors that were of importance and concern to

them. Parents examined single-item index cards with widely used parent-report measures of children's social-emotional behaviors and noted whether they felt the question was easy, difficult or uncomfortable to answer. They also participated in cognitive interviews to demonstrate their understanding of and rationale for answering items about their child's socio-emotional development.

Findings from this paper will highlight the themes that emerged across the three tasks and how perceptions of individual items varied by ethnic group. Discussion and implications will center on challenges and solutions to wording questions used in obtaining parent ratings of social-emotional development and ways researchers can begin to infuse these methods into their own survey work.

Tracking, Measuring, & Responding to Public Opinion

Multi-Level Policy Responsiveness to Public Opinion: From Statehouse to Street-Level.

Eric Plutzer, Penn State University (exp12@psu.edu) Michael B. Berkman, Penn State University (mabb1@psu.edu)

George Gallup understood that scientific public opinion polls could play an important role in democratic governance – both in instructing public officials about the public's preferences, and in monitoring government by highlighting instances in which government officials enact policies at odds with desires of ordinary citizens. For nearly a quarter century, political scientists and sociologists have investigated the correspondence between what people say they want (public opinion) and what they get (public policies). The association between public opinion and policy outputs has generally been found to be positive and modest in magnitude. Despite their wide range of units of analysis, analytic models, and substantive focus, these studies share a common trait: their dependent variable is public policy as enacted by legislatures or government agencies. We argue that for many issues, this does not go far enough; much is revealed by tracing policy all the way through to implementation. Theoretically, our model integrates responsiveness studies (which typically ignore implementation) with the literature on street-level bureaucracy (which typically ignores public opinion). Using public school curricular policy as an illustration of our approach, we show that public opinion exerts a powerful effect on policy implementation – a conclusion at odds with weak or null effects observed if one stops at written policy. We conclude that multi-level theories of policy responsiveness can reveal patterns of representation that are essential for a proper understanding of how governments translate mass demands into public policy.

Interpreting and Misinterpreting Trends in American Public Opinion

George Franklin Bishop, University of Cincinnati (george.bishop@uc.edu); Stephen T. Mockabee, University of Cincinnati (Stephen.Mockabee@uc.edu)

Tracking trends in the “state of the nation” has become a borderline obsession, with the instant availability of the “Gallup Daily Averages” (www.Gallup.com) and the “Flash Charts” at www.pollster.com, not to ignore the plethora of daily and weekly updates at www.pollingreport.com and the steady flow of reports from the Pew Research Center (<http://pewresearch.org>). Nor do many pollsters, pundits, and policy analysts hesitate to interpret such trends “on the spot” as if their meaning (and validity) were readily transparent. Obama's drop in approval ratings must mean this; the dip in consumer confidence must mean that; the shifts in “aggregate” party identification must obviously mean something else. It's just not that simple.

In this paper the authors demonstrate, empirically, with data from the past decade (2000-2009) in Gallup and other polls, that even with identically worded questions, the meaning of many well-established

indicators in national polls, such as presidential and congressional approval, the "mood of the country" and the like, is continuously shifting over time; that we are often comparing apples with oranges; and that such indicators can therefore be misleading. Furthermore, this type of systematic measurement error—shifts in the meaning-and-interpretation of identically worded questions over time—is not random and will, therefore, not cancel out in the aggregate. So the miracle-of-aggregation assumption that underpins much of the foundation for the research literature on collective abstractions such as the "rational public" and other macro models becomes highly questionable. As an alternative, we provide a method for modeling such systematic shifts in the meaning of identically worded items and incorporating these components into a more general multivariate model for explaining trends in public opinion.

Demographics and the Distribution of Local Political Knowledge

Lee Shaker, Princeton University (lee.shaker@gmail.com)

Knowledge of political affairs and processes amongst citizens is vital for a responsive and responsible democracy. A preponderance of evidence shows that knowledge at the national level is distributed unequally throughout the population in stable, predictable patterns in which people who are, in short, more socio-economically advantaged are likely to be more knowledgeable than others. There is very little research of who has local political knowledge, and preliminary results suggest that there are systematic divergences from the national-level contours. This paper uses survey data drawn from the city of Philadelphia to investigate the distribution of local political knowledge. The analyses reveal some surprising differences in the allocation of local political knowledge, suggesting that the SES-driven explanations of national political knowledge do not hold at the local level. These results cast new light on political scientists' understanding of citizens' political knowledge and, by extension, their political learning and competence.

The Unemployed American Worker in the Great Recession of 2009-2010: Economic, Social and Emotional Tolls.

Cliff Zukin, Rutgers University (zukin@rci.rutgers.edu); Jessica Godofsky, Rutgers University (jgodofsk@eden.rutgers.edu); Steven Fischer, Rutgers University (sfischer@rci.rutgers.edu); Carl Van Horn, Rutgers University (vanhorn@rci.rutgers.edu)

We present the results of a summer 2009 national survey of those who have lost their jobs in the great national recession. Using a web-based probability sample (Knowledge Network's Panel), 1,200 questionnaires were completed by those who had become unemployed in the 12 months prior to the survey, 900 of whom continued to be unemployed at the time of the survey. The original interviews with 1,200 was turned into a panel design, with a second wave of interviewing being conducted six months later (in February/March of 2010) to better understand the process of "reemployment" and the social, psychological and economic costs paid by the American worker along the way.

We describe: 1) Who the newly unemployed are and how they differ from long-term unemployed; 2) What services they need and want from government; 3) How they lost their jobs and job-seeking behaviors; 4) How people are coping economically, as well as emotionally and socially; 5) How many of those interviewed in August 2009 were still unemployed in March 2010; 6) What were the paths to reemployment: how do the newly reemployed compare to those still unemployed?

A Measured Threat: Factors Affecting Perceived Likelihood and Severity of National Threats.

John F. Kunz, ICF International (jkunz@icfi.com); Randall K. Thomas, ICF International (randall.k.thomas@gmail.com); Amy E. Falcone, ICF International (afalcone@icfi.com)

In the aftermath of 9/11 and Hurricane Katrina, voters and politicians in the U.S. became more aware of the vulnerabilities facing the nation from a number of man-made and natural disasters. Policy decisions are often made concerning both the perceived likelihood, however likely or unlikely, and the severity of the threat. As part of a larger web-based survey fielded through 2006 and 2007, respondents evaluated a series of 16 threats and were randomly assigned to assess each in terms of either their perceived likelihood or severity. The results were significantly affected by political party identification – Republicans considered ‘The banking system experiences a major financial collapse’ as having the lowest likelihood of all the events, and ‘A significant rise in the level of the oceans’ as having the lowest severity of all the events. Generally, Democrats indicated higher levels of perceived likelihood and severity for all events when compared to Republicans. Beyond political party and political ideology, we found that sex, age, region of country, and religiosity were all significant predictors of the various events. These results are integrated in the broader threat-perception literature and extend the literature into the area of perceived national threat.

Use of Web Surveys Across Diverse Populations

Web Questionnaire Usage in an Establishment Survey

Sarah J. Harris, RTI International (harris@rti.org); Paul P. Biemer, RTI International (ppb@rti.org); Michael A. Penne, RTI International (penne@rti.org)

In the last 25 years, the use of technology in the workplace has become more prominent across many industries. Workers increasingly depend on the Internet for communicating, researching, conferencing, and training. These trends in the workplace are also shaping the future of public opinion research as an increasing number of surveys provide respondents with the option to complete questionnaires via the Internet.

The purpose of this presentation is to explore the rates with which employees in specific industries have used an available Internet option for questionnaire completion on a long term national study. This study, begun in 2001, is designed to provide information on over 800 occupations in the United States. Employing a two-tiered random sampling method, phone representatives work with contacts at establishments across the country to identify and select a random sample of eligible employees within target occupations. The establishment contact then distributes survey packets that include both paper questionnaires and Web login information to those sampled employees. The employees are asked to complete the questionnaires during non-work times, using either the paper or Web option.

Our presentation will examine the rates with which employees across varying industries opt to complete an occupational questionnaire via the Internet rather than using the traditional paper and pencil version. We will present descriptive data exploring the correlations between the rates of Internet questionnaire completion and respondent education level, establishment size, and the use of technology in the workplace. By investigating metrics such as missing data, speed of questionnaire return, item level inconsistencies and outliers, we will also examine whether the mode of questionnaire completion has an effect on the quality of the data submitted. Based on our findings, we will make recommendations regarding the advantages and disadvantages of providing an Internet questionnaire option to respondents in an establishment survey.

An Experiment to Test the Feasibility and Quality of a Web-Based Questionnaire of Teachers

Robin Jacob, University of Michigan (rjacob@umich.edu); Lesli Scott, University of Michigan (ljscott@umich.edu); Brian Rowan, University of Michigan (browan@umich.edu)

In this study, a nationally representative stratified random sample of 877 first and fourth grade reading teachers were randomly assigned to one of two modes of questionnaire delivery (paper or web) to examine the relative costs and benefits of web versus mail administration (in terms of cost and time) and to examine how the mode of administering a teacher questionnaire affects (a) coverage error, (b) non-response error, and (c) measurement error in the context of educational research. Web surveys have the potential to achieve economies of scale, allowing greater sample size within a fixed budget. However, concerns about the logistics and quality of collecting data via the web have made many educational researchers wary of switching modes. Initial results from this study indicate that these concerns may be warranted. While email addresses were obtained from all but 4% of the sample and mailing addresses were obtained for the entire sample, approximately 12% of the emails obtained were undeliverable even after attempts to correct email addresses were made. After three contacts, the response rate for the web group was only 43% while the response rate for the mail group was 59%⁹ (based on AAPOR response rate 1). Individual item non-response was also higher among the web respondents. The problem was most pronounced for items that asked individuals to “mark all that apply” or that were near the end of the questionnaire. Differences in the characteristics of responders and non-responders will be explored. Initial findings also suggest that switching modes may introduce measurement error. While the costs for the web survey were approximately 40% lower than for the mail survey, the time it took respondents to return their surveys did not differ substantially for the two groups.

Web Surveys when Databases have Incomplete Email Records: The Impact of Modifying the Method of Respondent Contact on Web Survey Response Rates.

Michael G. Elasmr, Boston University (elasmar@bu.edu)

What happens when you need to do a web survey but your database of email records is incomplete? This paper reports the findings of a 2009 study of university alumni where a web survey was utilized. In the database from which the sample was drawn, approximately 40% of cases had invalid and/or questionable email records. In order to ensure that the methodology adopted represents all alumni in the database, a decision was made to recruit participants differently depending on the type of email record that existed for them. Recruitment, in this context, consisted of a letter and several reminders inviting those who were sampled to access a unique URL and take a web survey about their experience with the academic program that they had attended. Those who had an updated email record were contacted by email only (60%); those who had a questionable email record were contacted by email and by U.S. mail (28%); and those who had no email record were contacted by U.S. mail only (12%). A total of 1138 invitations were sent out. Given the use of a mixed invitation mode, the survey field period lasted 30 days. The web survey response rates varied greatly by recruitment method: A response rate of 58% was achieved for email only invitations, a response rate of 41% was achieved for email and U.S. mail invitations, and a response rate of 12% was achieved for U.S. mail only invitations. Specific methodological details are provided in the paper and implications are drawn for future studies wishing to use a web survey method and a database of incomplete email records.

College Road Trip: Transforming the NHTS into a Web-Based Travel Diary Survey of University Students in Virginia.

Thomas M. Guterbock, Center for Survey Research, University of Virginia (tmg1p@virginia.edu); Tancy Vandecar-Burdin, Old Dominion University (tvandeca@odu.edu); Susan G. White, Virginia Commonwealth University (whitesg@vcu.edu); Susan Willis-Walton, Virginia Tech (suwillis@vt.edu); Paul Agnello, Virginia Department of Transportation (Paul.Agnello@VDOT.Virginia.gov)

In the National Household Travel Survey (NHTS) adults are assigned to a 'travel day,' recruited by phone, mailed paper travel diaries, and then debriefed by telephone to obtain detailed trip information for all household members. Because the method relies on landline RDD sampling, it typically misses college students, most of whom are cell phone-only users. In order to refine travel demand models for localities containing colleges, VDOT contracted in 2009 with ODU, UVa, VCU and Virginia Tech to survey their students. Although the task of reporting trip information from a travel diary presents significant cognitive demands, the survey was to be conducted via the Internet.

This assignment presented challenges in several dimensions:

- Conceptualization. The traditional typology of trip purposes did not fit.
- Travel modes. Walking and bike trips are much more frequent for students.
- Unit of analysis. Households or individuals?
- Reachability. Listings of student postal addresses and phone numbers were unreliable.
- Travel day assignment. The universities used different methods for assigning and re-assigning travel days.
- Motivation. What incentives to use?
- Diary visualization. Our on-screen diary could not match the format of the NHTS-style diary we mailed out.
- Capture of travel locations. Could students provide codable destination information?
- Implementation. The study required staff action seven days a week.
- Quality of measurement. With no interviewer present to guide the response process, would students recall and report correctly?

We ultimately succeeded in collecting a rich set of travel data that will support travel demand modeling with campus populations. However, between 15 and 20 percent of students made detectable errors in reporting their trips. Based on respondent debriefings and analysis of the problem cases, we are able to recommend improvements that could streamline implementation and reduce the error rate in future student travel surveys.

Experimental Tests on Response Rates in Student Web Surveys: What Works and What Doesn't.
Joel D. Bloom, The University at Albany, SUNY (joeldbloom@gmail.com)

In this paper, I present results from a total of 6 response rate experiments conducted during the administration of three separate student surveys at The University at Albany, SUNY, a medium-sized research university in the Northeast. The most important finding is that in two experimental treatments, personalizing the invitation or reminder e-mails (e.g., "Dear John") produced a roughly 25% increase in the raw number of responses compared to the control group that received an anonymous invitation (e.g., "Dear Student"). This confirms many years of findings related to personalization of mail survey solicitations. Other treatments produced no increase in response rate. These included changes in the subject line to either mention the word "survey" or not, as well as three experiments related to e-mail pre-notifications. In none of these treatments did any version of the e-mail pre-notification (personalized vs. anonymous; from the Vice Provost's account vs. from the UAlbany Survey e-mail account; pre-notification vs. no pre-notification) have any impact on response rates compared to the group that received no pre-notification.

AAPOR Concurrent Session C

Friday, May 14, 2010

1:45 p.m. - 3:15 p.m.

Alternative Data Collection Techniques: Biospecimens, GPS & Second Life.

Factors Affecting Participation in the Collection of Biological Samples for Genetic Testing in a Survey of Detroit Residents

Mark Morgan, Abt SRBI (m.morgan@srbi.com); Christine Cowles, Abt SRBI (c.cowles@srbi.com)

The collection of biological specimens is increasing the usefulness of survey data to scientific inquiry around mental health. Lab testing for specific DNA or RNA sequences can greatly enhance the understanding of biological correlates to observed health symptoms. Together with survey data on self reported behaviors this can expand the opportunities for researchers to examine interactions between multiple possible causes for disease. A high level of participation in the biological sample phase of studies is critical to the success of these research efforts.

The Detroit Neighborhood Health Study (DNHS) is a longitudinal research project among a household sample of residents in Detroit. At baseline, 1,500 respondents were surveyed by telephone and invited to provide blood samples after the survey. Biological samples were collected at baseline and are also being collected in Year 2.

An exploration of the issues that contribute to and detract from the success of biological sample collection in this research will be examined to help contribute to best practices recommendations in this growing sub-field within survey research.

Implementing Biospecimen Collection and Tracking Methods on a Large-Scale Field Study

Amy Hendershott, RTI International (hendershott@rti.org); Stacey Weger, RTI International (sweger@rti.org); Gina Kilpatrick, RTI International (ginak@rti.org)

In studies where blood and saliva biospecimen samples are collected by multiple field interviewers for analysis, standard guidelines must be implemented to ensure consistent collection, storage, and shipment. With a large number of biospecimen samples being obtained daily, it is also critical to track the collections as they are sent from the field to the labs that receive and process the samples so that all collected samples are closely monitored and accounted for.

For a recent longitudinal study conducted by RTI, detailed biospecimen collection protocols were developed to ensure the acquisition of high quality samples. Field interviewers were required to complete an extensive training on biospecimen collection, including a certification on collection, storage and shipment procedures. In cases where respondents required a phone interview rather than home visit, a special process was implemented to allow them to receive a saliva collection kit. These respondents were provided with thorough instructions, and were asked to self-collect a saliva sample and ship the collection kit back to the processing lab. As a means for accurately tracking and reporting biospecimen collection and shipment data, an intricate Biospecimen Tracking System (BTS) was developed. This enabled project staff and the labs to enter information about the collected samples, and provide data quality

reports on the collections. Feedback regarding sample quality was provided to project management by the processing labs, and was immediately shared with field interviewers. Improvement methods, such as one-on-one retraining, were implemented as needed. In an effort to share the knowledge and lessons learned, this presentation describes the innovative, non-invasive biospecimen collection methods and tracking methods used for this study, as well as biospecimen collection success rates based on the implementation of improvement methods.

Reporting Travel by GPS Tracking vs. Diary Self Reporting

Mindy Rhindress, Abt SRBI (m.rhindress@srbi.com); Laurie Wargelin, Abt SRBI (l.wargelin@srbi.com)

This paper reports the results of a pioneering Household Travel Survey using Geographic Position System (GPS) devices in place of traditional paper diaries.

Traditionally, large-scale Household Travel Surveys have been conducted by assigning randomly selected households a travel day and mailing out travel diaries for each member of the household. Respondents self-report their travel by phone, web, or mail-in. As in other population surveys, non-response bias is corrected to the extent possible using iterative proportional fitting to match census-based population targets. However, for HTSs, “item non-response” is more problematic; particularly the challenge of missing trips. People neglect to report trips they made for any number of reasons including not placing importance on reporting short trips or stops on the way to somewhere else, or simply due to respondent burden—they tire of the redundant process of self-reporting name, address, and activity for each location.

Small subsample efforts have been undertaken to compare trips recorded by portable GPS devices with self-reported trip rates. Based on results, the imputed missing diary vehicle trip rate was 17.8% of all GPS trips.

Given these findings, US DOT and Ohio DOT recently funded the first large-scale GPS (only) HTS in the country. No diaries are being assigned. Every member of 6,000 recruited households, twelve years and older, is being asked to carry a personal, pocket size GPS unit, which will record their travel over three assigned days. Respondents are asked to verify their GPS recorded travel via a web-based survey where a Google map image of their travel for each day is displayed.

We will report findings from the first interim report of 500 completed households for the Ohio-based GPS-only survey, comparing non-response bias and item response bias (particularly missing trips) with the 2006 Puget Sound Regional Council sample of 4,000 household diary self reports.

Insights into Population Segmentation Using the InfAlign Map

David Fan, University of Minnesota (dfan@umn.edu); Kenneth Blake, Middle Tennessee State University (kblake@mtsu.edu); Jason Reineke, Middle Tennessee State University (jbreineke@gmail.com); Robert Wyatt, Middle Tennessee State University (robertowatt@gmail.com)

One common goal of surveys is to segment the population into subpopulations with common characteristics. This paper shows the utility of the InfAlign map for this purpose using data from a recent poll of attitudes toward health care reform.

For the map, survey data are written as a table of numbers with responses in the rows and respondents in the columns. Each number is converted to a vertical bar graph so that all responses for a respondent become a vertical stack of bars. The map is drawn on a computer screen with a vertical cursor bar that can move over different bar stacks. When the cursor is over a stack, the ID and other information about the respondent are shown. This information allows for removal of the labels for the bars thereby permitting the bars to shrink in width.

As with maps in general, the InfAlign map can be zoomed out to lower resolution by displaying responses for just 300 random respondents, a number that can be seen on a single computer screen. The map can also be zoomed in by showing only respondents meeting selection criteria. A further zoom is the mapping of a subset of all responses.

Automobile road maps do not specify routes; instead, the user decides the desired path. In analogy, the InfAlign map also provides no inferences; the user consults the map to form hypotheses to be confirmed by other procedures.

In contrast, methods such as the Chi-squared Automatic Interaction Detector (CHAID) provide relationships between survey responses directly. That is like GPS devices prescribing a travel plan.

The InfAlign map showed that undecided's and supporters for health care reform could be merged into a common subpopulation given their similarities in both ideology and support for the public option. These results complemented those from CHAID.

“Can Survey Research have a Second Life? The Opportunities and Challenges in Bringing Methodological Rigor to Virtual Worlds Research.” Kelly Foster, University of Georgia (knfoster@uga.edu)

The recent expansion of technologies - particularly in cell phone usage and internet connectivity - have brought both feast and famine to the survey research and public opinion field. The difficulty in collecting data by traditional means has challenged researchers to diversify their methods. Those who have had success with cell phone and web surveying are now tentatively exploring the field of virtual world's research where one is only limited by his/her imagination. Researchers can test interviewer effects by modifying the apparent gender, weight, height or skin coloring of the interviewer at the touch of a button. Data collection can be done through focus groups, in-depth interviews, in-person surveys, and CAPI-like surveys - often at a fraction of the cost of conducting a similar survey in real life (Dean et al. 2009).

However, the methodological foundation supporting research in virtual worlds is complicated because few, if any, worlds openly share information about their users. Unfortunately, this makes it all but impossible to design a random sample where everyone has an equal probability of selection. At present, most research in virtual worlds relies on convenience samples. In fact, there is only one survey completed in Second Life (one of the largest and most active virtual worlds) that can make a reasonable claim to have representative data (Bell, Castronova et al. 2009).

The author seeks to add to this body of literature by exploring the methodological issues researchers face when conducting research in virtual worlds. In addition, the author will present preliminary data comparing Second Life residents to a nationally representative sample on demographic characteristics as well as health related attitudes and behaviors. It is the author's intent to provide an overview of the methods used, what worked and did not work, and suggestions for future sampling in virtual worlds.

Challenges of Language Translation

Testing the Effects of a Multi-Lingual Brochure in the American Community Survey.

Megha Joshipura, U.S. Census Bureau (megha.p.joshipura@census.gov)

The American Community Survey (ACS) questionnaire currently includes a message on the cover that informs respondents how they can obtain assistance in English or Spanish. However, no messages currently exist on the ACS questionnaire or any other ACS mailing piece that explain how households that require assistance in other languages can receive assistance. Therefore, the ACS telephone and personal visit follow-up operations are primarily responsible for data collection from these populations. We are interested in trying to increase the number of interviews conducted for non-English speaking households prior to telephone and personal visit follow-up activities, thus providing a less expensive mode for collecting data from these households. Moving respondents into the mail or telephone mode of data collection also improves the reliability of ACS estimates for speakers of these languages since personal visit cases are subsampled. In particular, we would like to use the mail mode to explain to Spanish, Chinese, Korean, and Russian households that they can receive assistance in these languages. To do so, we developed and cognitively tested a multi-lingual brochure to be included and tested in various ACS mail packages. The brochure includes instructions on how households can obtain telephone assistance in the language they speak, and provides some additional ACS information to give context. The brochure includes Spanish, Chinese, Korean, and Russian translations, with English provided as a reference. This paper examines the potential effects of the multi-lingual brochure on the ACS response.

Designing Survey Materials to Target the Hispanic Population

Lorelle Vanno, The Nielsen Company (lorelle.vanno@nielsen.com); Teresa (Ye) Jin, The Nielsen Company (teresa.jin@nielsen.com); TraShawna Boals, The Nielsen Company (trashawna.boals@nielsen.com); Jennie Lai, The Nielsen Company (jennie.lai@nielsen.com); Charles D Shuttles, The Nielsen Company (chuck.shuttles@nielsen.com); Michael Link, The Nielsen Company (michael.link@nielsen.com)

Targeting hard-to-reach demographics has become a standard in survey research. The growth of the Hispanic US population has an ever-increasing impact on national and regional survey estimates, as sometimes estimates cannot accurately represent the Hispanic population in a survey universe. Often we target hard-to-reach demographics like the Hispanic population in the US by providing higher incentives. However, what if there was a way that we could target the Hispanic demographic by customizing our bilingual materials to target Spanish-dominant Hispanics specifically and therefore increasing response rates for this demographic without increasing incentives?

In order to test the hypothesis that improving bilingual materials to target the Hispanic population could increase response rates, The Nielsen Company collaborated with a Hispanic Marketing Firm to improve the bilingual materials for the Radio diary pre-recruitment survey. The pre-recruitment survey is the process Nielsen uses to collect phone numbers from the unlisted portion of the Address Based Sample. Aspects of the Hispanic culture were incorporated into the text, graphics, and images of the materials; these changes included focusing on the family and community and enjoying music. Instead of simply translating English text into Spanish, the decentering translation technique was used to ensure the verbiage for both English and Spanish was culturally appropriate.

To prove or disprove the hypothesis, response rates for the questionnaire pre and post re-design will be analyzed. This will include the response rates and the percentage of returning households that are

Spanish-dominant Hispanics. If successful, in the future, this could lead to higher response rates allowing for a decrease in the additional sample needed for the Hispanic population and possibly even reducing the amount of the incentive that is currently used to target this population. Both of these options would reduce the cost of the survey.

Scoring English Language Assessments for Speakers of Other Languages: Guidelines, Training and Quality Assurance

Jane Shore, Educational Testing Service (jshore@ets.org); Susan Sprachman, Mathematica Policy Research (ssprachman@mathematica-mpr.com); Jennifer Lentini, Educational Testing Service (jlentini@ets.org); Kathy Buek, Mathematica Policy Research (kbuek@mathematica-mpr.com)

Survey research projects increasingly include assessments of language and literacy. Many assessments require that test administrators be sensitive to the linguistic diversity of test takers. None, however, offer administration and scoring guidelines for use in real-time by testers. This leaves implementation of the guidelines open to interpretation. This paper will discuss the guidelines for linguistically sensitive scoring, implementation of the guidelines, and quality assurance procedures.

The issue of linguistic sensitivity becomes especially pertinent when tests are used with respondents from diverse linguistic backgrounds. This paper discusses the development of linguistically sensitive guidelines for scoring alphabet naming, word identification and decoding tests that require oral responses. It describes how these guidelines were implemented in-real time by field interviewers/testers in a study of low literacy adult English learners. It also addresses assessor reliability, training and quality assurance procedures. It concludes with lessons learned that can be generalized to other testing conducted as part of survey research projects.

The study sample included native speakers of Spanish, Haitian Creole, Chinese, and Armenian, and other languages in five programs for adult English learners. To address the differences in pronunciation for each group, linguists at ETS developed scoring guidelines for each of the major language groups in the study. For each test item, the guidelines specified the various pronunciations of answers that were to be scored correct or incorrect.

Assessors were bilingual. To the extent possible Mathematica matched assessors with students of the same linguistic background. Assessors were trained to implement the scoring guidelines for the students in their sample. Training included “rules” underlying the guidelines, and practice scoring sessions using audio-recorded tests with ESL students. During data collection, Mathematica and ETS conducted different levels of reviews of audio-recorded assessments to maintain data quality. Mathematica delivered individualized feedback and re-training to assessors as needed.

Agreement Answer Scale Design For Multilingual Surveys: Effects Of Translation-Related Changes In Verbal Labels on Response Styles and Response Distributions.

Ana Villar, Stanford University (villar@stanford.edu)

Answer scales in survey instruments are widely used, but little is known about how to choose verbal descriptors as labels. In multilingual research, this matter is further complicated because answer scales must be appropriate for all languages and function comparatively. Comparing source questionnaires to translations of multinational projects, it was observed that certain verbal features differed across languages, countries, and modules. This paper empirically investigates the effect of such changes on response distributions. The verbal feature examined is the presence or absence of an intensity modifier in

the second and fourth labels of a 5-point agreement scale: strongly agree, (somewhat) agree, neither/nor, (somewhat) disagree, strongly disagree.

Data from more than 40 countries of the International Social Survey Programme were analyzed. It was expected that Using multilevel models in five ISSP modules, the analyses show that, as predicted, adding an intensity modifier increases the tendency to select the endpoints.

In a second study, Graded Response Models were used to compare each of the response categories of the two answer scale versions. For the agreement answer scale under study, adding an intensity modifier made the scale points less useful to measure the underlying attitude than were scale points without modifier.

These findings suggest that modifications made to answer scale language versions are a critical source of variability in response patterns and distributions.

Developing a Systematic Process for Translation Expert Review: The Translation Appraisal System (TAS-10)

M. Mandy Sha, RTI International (msha@rti.org); Hyunjoo Park, RTI International (mpark@rti.org); Yuling Pan, U.S. Census Bureau (yuling.pan@census.gov)

Several approaches can be used to assess the quality of a translated instrument. The most cited method is Back Translation; however, recent literature recommends conducting an expert review because the results are more content relevant and there are fewer false positives among the identified problems (Harkness et al., 2009). Yet, formal procedures for Expert Review have not been well-documented. This paper describes research conducted to develop a systematic appraisal of translation by language experts.

This study is based on a translation review and cognitive testing project undertaken at the U.S. Census Bureau to pretest the Chinese and Korean translation of the American Community Survey (ACS) Language Assistance Guide (LAG). The translation covered a variety of topics, such as household composition, housing, income, and health.

The Translation Appraisal System (TAS) is a coding tool designed to assist the translation Expert Review process. Applying the concept of documenting question features by using an item taxonomy of the cognitive demands of a question (Dean et al., 2007), TAS is geared toward translated items and the common sources of translation problems. It guides language experts to complete translation Expert Review in a systematic way. We also conducted an experiment to examine the utility of the TAS to the Expert Review method. Four Chinese language experts and four Korean language experts assessed the translation of the ACS LAG and were assigned by experience (as defined by years of experience in translation and pretesting). The less experienced language experts used the TAS while the more experienced did not. Our hypothesis is that when aided with TAS, less experienced language experts can do as good of a job of Expert Review or even better than more experienced language experts. The number, type, and relevancy of the translation problems as well as time taken are examined.

Consumer & Media-Related Attitudes

Video Consumer Mapping Direct Observation Study: Out of Home Video Habits of U.S. Adults.

Norman Trussell, The Nielsen Company (norman.trussell@nielsen.com); Michael Link, The Nielsen Company (michael.link@nielsen.com)

The Video Consumer Mapping Study (VCM), the largest and most extensive media use observational study ever conducted, explores U.S. consumers' exposure to various forms of media, with the primary focus on television and video. Other media consumption studies have been limited to comparing behaviors among datasets from multiple surveys for various media sources. The "single source" methodology used in this study is a uniquely powerful approach in that media consumption behaviors are assessed across an array of media sources for the same set of individuals. This landmark research project was sponsored by the Council for Research Excellence (CRE), with funding from The Nielsen Company. Using a direct observation method, the study gathered data about current video consumption to serve as a guide to future audience measurement efforts. VCM was conducted among participants in five DMAs: Atlanta, Chicago, Dallas Philadelphia, and Seattle. The sample included 752 observed days (376 participants with half in spring 2008 and half in fall 2008), and over ¾ million minutes of observation, at 10 second resolutions throughout those days. A wealth of data was collected from this study, but the focus here is on out of home video viewing/exposure behavior. A brief review of the methodology and limitations of the study will be presented. Key metrics from this study including TV viewing at home, in other people's home, at work, in cars/public transportation, and in other locations will be reported. These viewing levels will be examined by genre of programming and live versus time-shifted viewing via a digital video recorder (DVR). Differences in viewing levels will be presented by various demographic groups including age, gender, race/ethnicity, income, employment status, marital status, education and presence of children, both overall and by location. Findings from this study are compared with those from others on out of home viewing.

Following the Users of Online Social Networks: An Observational Study of Their Life Activities and Media Use

Jennie Lai, The Nielsen Company (Jennie.Lai@Nielsen.com); Michael Link, The Nielsen Company (Michael.Link@Nielsen.com)

Given the exponential growth of the social network community on the Internet, there has been much reported on their behavior online but what about their behavior offline, in particular on their everyday activities and other forms of media exposure? The online activities of this group can be measured with minimal burden on the respondents through Internet-mediated method; however, their offline activities can be difficult to measure accurately if using self-reporting or interviewer-administered instruments. In an effort to understand consumer exposure to multiple media platforms, The Council for Research Excellence sponsored the largest and most extensive observational studies ever conducted to measure media exposure in the constantly evolving multimedia landscape. This large-scale study was executed by Ball State University and Sequent Partners to conduct real-time observation by trained observers using a computer-assisted data entry device of two full waking days of each participant once in Spring 2008 then again in Fall 2008. The final sample included 495 participants, 952 observed days, over ¾ million minutes of observation at 10 second resolution throughout those days. The analysis of this research will focus on the 69 participants used online social networks and their life activities along with media usage during the days observed. The granularity of the data provided a high quality view of the day-in-the-life of these participants with unprecedented level of real-time data but without the self-reporting or recall bias. Using observational method, this groundbreaking study contributes to our understanding on where this emerging cohort spent most of their time, what life activities they partook and the various forms of media outlet they were exposed in addition to their demographic profile. The insights shared in this in-depth qualitative study can help survey researchers' gain greater understanding of this community of online social network users and their activities offline.

The Relationship Between News Preference and News Consumption: A Comparison Between News Preference, a Metered Measure of News Viewing, Self-Reports of News Consumption.

Su Jung Kim, Northwestern University (ksio2@northwestern.edu)

The relationship between attitude and behavior has long been a subject of academic investigation in social sciences. Among media effect scholars, there have been many concerns about using news preference as a predictor or proxy of news consumption due to the discrepancy between news preference and an actual level of news viewing caused by the social desirability bias that people try to give a positive impression about themselves to an interviewer or by respondents' inability to accurately recall their level of news consumption. This study attempts to investigate this controversial relationship between news preference and news consumption by using a merged survey dataset collected from 1,500 individuals of the household peplemeter panel of TNS Media Korea. The mixed-mode data collection (i.e. electronically recorded television viewing log and a telephone survey conducted to the peplemeter panel) provides a unique opportunity to examine the relationship between news preference, a metered measure of network news viewing, and self-reports of other news media use such as newspaper, cable television, and the Internet.

The results show that news preference and actual news consumption (i.e. a metered measure of news viewing) was positively correlated, but the association was very weak ($r = .081$, $p < .01$). The positive relationship between news preference and self-reports of other news media use was also found at a stronger degree than that of news preference and actual news viewing (for newspaper, $r = .156$, for cable news $r = .212$, for online news, $r = .132$, $p < .001$). The findings were consistent across key demographic and socioeconomic variables (i.e. gender, age, education, income). The results suggest that participant's perceived level of news preference can be used as a predictor of self-reports of news media use, but researchers should caution when they consider news preference as a predictor or a proxy of an actual level of news consumption.

Bowling Together in the Internet Era: A Re-Examination of the Putnam Hypothesis

Jon D. Miller, Michigan State University (jdmiller@msu.edu); Linda G Kimmel, Michigan State University (kimmell2@msu.edu); Marian Mitchell, Michigan State University (mitch285@msu.edu)

Robert Putnam's *Bowling Alone* (2000) pointed to fundamental changes in the social interactions of American adults over the last 50 years. He looked at group memberships and participation in various kinds of community activities and organizations and concluded that American adults were becoming markedly less social and more isolated. Symbolized by mid-week bowling leagues, Putnam concluded that a combination of work, television, and other factors have combined to eliminate our desire to socialize in these traditional forms. Putnam briefly examines the impact of the Internet, but his data was from the early 1990's and he found only a minimal impact. This paper uses a combination of time series national measurements of media use and social interaction and a detailed 2007-8 panel study to explore how adults interact and socialize in the Internet Era. The 2007-8 panel was designed to explore some of Katz and Lazarsfeld's (1955) original hypotheses about personal influence and the proposed paper is an extension of that work. All of the respondents ($N = 1000$) have been geo-coded and the analysis will look at central city, suburban, and non-metropolitan locations as factors as well as extensive sets of measures of media use, Internet use, personal conversation and visiting, and other kinds of subject-specific interactions (trading recipes, political campaigns, and health issues). The analysis will utilize a set of structural equation models, but the results will be presented in a manner that will make the results accessible to audience members with and without an understanding of these techniques.

Exploring Media Use across the Non-Western World.

Joachim Bruess, InterMedia (bruessj@intermedia.org)

During the past two decades not only the Western but even more so the non-Western world has seen an enormous increase in access to new Information and Communication Technologies (ICT). Especially mobile technology has made it increasingly easy to get news and information across many countries in the world. However, access to ICT does not necessarily mean that more and more people use it. Depending on costs for instance, modern ICT might be restricted to the wealthier segments of societies. Well established and formal gender relations might primarily include men in some countries whereas women are less able to use ICT in order to get news and information. Also, different age groups might prefer different means of getting news and information. Finally, given the substantial variation across non-Western countries – consider for instance wealth, population size and education - it seems timely to explore the so-called digital divide across non-Western countries.

To address these questions, this paper presents an analysis of an audience based Media Use Index (MUI). Based on large, random, representative samples from more than 60 countries - typically involving more than 2000 respondents per country - the analysis combines more traditional (newspapers, radio and TV) with modern means (mobile phone, internet) of getting news and information. Countries will be ranked according to their media use and they will be compared to their peers based on their media use. Subgroup analysis will explore gaps related to socialization, diffusion and markets.

The value of the MUI is that it facilitates a comparison of a use measure with access data. In addition, a cross country comparison might help improve investment strategies for development initiatives. Within country results will highlight differences that lead to further specification in investment and training.

Improving Questionnaire Design to Reduce Bias

Measuring Attitudes: Using Branching and Numerical Scales

Davide Lubian, Sapienza - University of Rome (ITALY) (davide.lubian@uniroma1.it)

Past research has recognized the effectiveness of verbally labeled branching questions with two aims: to simplify the judgment task, and to explore and optimize the midpoint answer. However, I recommend branching since positive and negative extremity answers are asked in the second phase, both with positive labels, resulting in the same unit of measurement.

Numerical scales are usually inappropriate for measuring attitudes because they exclude the negative part of the continuum. Accordingly, I conducted a split-ballot experiment (SAQ; N=146) to study the effect of branching with numerical scales, comparing 1-10 scales of trust toward 16 different institutions with their branching versions (i.e. dichotomous question followed by 5-point scale of extremity). Only a few studies have attempted to disambiguate the meaning of numerical scales using negative numbers. Their findings showed a positivity bias effect, which significantly shifts distributions to the positive side. I argue that there is also an avoidance effect with negative numbers, because respondents do not use them in everyday life. The solution that includes negative answers, but not negative numbers with rating scales, is branching.

Results of an MTMM analysis, including four Guttman scales, showed higher convergent and divergent validity for the branching format. Although branching scales had an average mean higher than the unipolar scales (5.6 vs. 4.6), respondents did not avoid negative answers (46% in the first step) as happened elsewhere with negative numbers. Furthermore, branching allows a better use of the scale gradation (SD 2.7 vs. 2.4), especially for the positive side.

This study warns researchers against employing unipolar rating scales for measuring bipolar constructs, as attitudes are, suggesting that negative answers, either verbally or numerically labeled, have a different unit of measurement than positive answers. This asymmetry implies a comparison problem between the two directions of the evaluative continuum, which branching can reasonably resolve.

Response Behaviors in Self-Administered Multiple Questionnaire Surveys

Jill N Lacey, US Government Accountability Office (laceyj@gao.gov); Carl M Ramirez, U.S. Government Accountability Office (ramirezcc@gao.gov)

Survey respondents are sometimes asked to complete multiple questionnaires for multiple instances that are not enumerable by the researcher when a sample is designed. That is, the reporting units (individual respondents) are pre-identifiable and can be sampled, but the number and identities of the units of analysis (household members, subsidiary business entities, etc.) are only discovered during the survey itself. This condition presents questionnaire design challenges – the existing literature concerning one such design choice is reviewed: the topic-based approach (one question presented to the respondent for every unit of analysis simultaneously, often resulting in a grid format) versus a serial, unit-based approach (all questions presented to the respondent for one unit of analysis at a time). One extreme version of the serial approach is asking the respondent to complete multiple copies of a standalone questionnaire for each unit of analysis. This paper examines the behavior of respondents faced with such requests. Using paradata from GAO web surveys, the navigational patterns of within- and across-questionnaire activity are analyzed. Navigational patterns are varied and complex, and our research suggests that in some multiple-questionnaire surveys, the level of respondent effort may be very high – a substantial number of respondents approach such tasks by jumping back and forth between copies of instruments on a question-by-question basis, for example. We conclude with implications for survey burden and quality.

Is Response Heaping a Form of Survey Satisficing?

Allyson L Holbrook, Survey Research Laboratory, University of Illinois at Chicago (allyson@uic.edu); Sowmya Anand, Survey Research Laboratory, University of Illinois (sowmya@srl.uic.edu)

Response heaping occurs when respondents choose common numbers such as 5 or 10 at higher rates than would be expected (i.e., at higher rates than noncommon numbers such as 4, 6, or 9). Researchers have suggested that response heaping may reflect satisficing processes whereby respondents do not fully and carefully going through the cognitive steps necessary to generate the most accurate response (Schaeffer and Presser 2003). We explored whether response heaping reflects satisficing processes by examining whether response heaping occurs more for respondents most likely to satisfice (e.g., those who have fewer cognitive skills and are therefore less able to go through the cognitive steps necessary to optimize) and under conditions most likely to foster satisficing (e.g., later in a survey or questionnaire when respondents are fatigued).

References:

Schaeffer, N. C. and Presser, S. (2003). The science of asking questions. *Annual Review of Sociology*, 29, 65-88.

Communal Belief and Calibrated Attendance: New Measurement Strategies for Contextualizing Religious Belief and Behavior.

Robert P. Jones, Public Religion Research (rjones@publicreligion.org); Daniel Cox, Public Religion Research (dcox@publicreligion.org)

As public opinion research on religion has grown over the last decade, survey researchers have continued to refine current measures of affiliation, belief and practice. In 2006, a group of political scientists introduced several new items in the 2006 American National Election Study that asked about social religious practices. These measures were introduced to complement current items like frequency of worship attendance and prayer that measured individual piety and were known to strongly correlate with partisanship and vote choice. The primary argument for including these measures was that “the social and communal aspects of religion that are present in the daily lives of people of faith from different religious traditions are often poorly tapped.”

This paper builds on these efforts by testing new measures of religious and moral belief that can be used to analyze individualistic versus communal orientations. We will include four items in a statewide survey of 3,000 Californians that will be conducted at the end of 2009. These items were originally included in a survey of mainline Protestant clergy conducted by Public Religion Research and the Bliss Institute for Applied Politics at the University of Akron.

Additionally, we will include a new measure of religious attendance that asks respondents who attend religious services how many services their place of worship offers. Because so many religious traditions do not offer more than one service per week, this variable will allow for a more accurate picture of religious participation in the United States, appropriately sensitive to differences between traditions.

The Effectiveness of an Interactive Web Survey in Decreasing Satisficing and Social Desirability Bias

Gina K Walejko, Northwestern University (gina@u.northwestern.edu)

A myriad of studies compare the estimates of identical questionnaires between different modes in order to determine patterns in survey measurement error. For example, some researchers compare estimates of socially undesirable behaviors between self-administered questionnaires like pen-and-paper or web and interviewer modes like telephone or face-to-face. Other studies compare proxies of satisficing. Such research influences our definitions of an “interviewer mode” and a “self-administered mode.”

This study examines how taken-for-granted characteristics of survey modes blur when self-administered web surveys become interactive. Using a 294-person survey experiment, we compare identical questionnaires between three modes: face-to-face interviewer, traditional web, and experimental web. The experimental web condition uses a survey programmed to slow respondents down, re-question respondents who appear to answer with socially desirable responses, and probe respondents for more responses to open questions. This condition is then compared to an identical face-to-face survey that uses probes and re-questions respondents and a “traditional” web survey that does not use such interactions.

Experimental results show intriguing albeit mixed results. On certain questions, web respondents are less likely to answer in a socially desirable fashion than face-to-face respondents, regardless of the interactive nature of the web survey. For these questions, mode matters more than the interaction. For other questions, the interactive web mode and face-to-face mode provide identical response patterns, illustrating how interactions such as probes and requesting can lead to higher data quality. The same modal pattern emerges for issues of satisficing.

These discrepancies in results illustrate how, as gender- and race-of-interviewer effects manifest only on specific questions, so do interactions on self-administered questionnaires matter for data quality on specific questions and sets of questions. The study concludes that web surveys should be viewed as more than static, self-administered questionnaires while acknowledging that survey mode boundaries are blurry not discrete.

Improving Response Through Incentives

All Things Incentive: Exploring the Best Combination of Incentive Conditions. Erin B Foster, NORC at the University of Chicago (foster-erin@norc.org); Alicia M Frasier, NORC at the University of Chicago (frasier-alicia@norc.org); Heather M Morrison, NORC at the University of Chicago (morrison-heather@norc.org); Kathleen S O'Connor, National Center for Health Statistics, CDC (kdo7@cdc.gov); Stephen J Blumberg, National Center for Health Statistics, CDC (swb5@cdc.gov)

Many surveys have successfully increased response rates by using monetary incentives. To make these incentive efforts cost-effective, survey researchers seek to balance the benefits to survey quality that may come with increased response rates against the costs of implementing an incentive effort. To achieve this balance is no small task given that identifying the maximally efficient incentive effort for a particular survey requires researchers to take into account a multitude of factors, including effective timing of the incentive offer, the value of the offer, and how the incentive offer should be made - prepaid or promised. For prepaid incentives researchers must also consider mail mode (first-class or priority).

To identify the maximally efficient incentive approach for the 2009-10 National Survey of Children with Special Health Care Needs we developed an incentive experiment to investigate each of these elements. The 2009-10 NS-CSHCN, is conducted through the State and Local Area Integrated Telephone Survey (SLAITS) mechanism of the National Center for Health Statistics on behalf of the Maternal and Child Health Bureau under contract with NORC at the University of Chicago. A screening interview is conducted for all households with children to determine each child's special needs status. Demographic information is also collected on each screened child.

We will analyze data from an experiment designed to identify the most effective approach to incentive-offer timing (based upon known eligibility, after 1 or 2 refusals, or not at all), incentive type (promised versus prepaid), mailing type (priority versus first class), and incentive value (\$1, \$5, \$10, \$11, \$15) for the 2009-10 NS-CSHCN. To achieve this analytic complexity, our incentive experiment contains 8 different experiment groups. These elements will be examined individually and collectively for impact on number of completed interviews, response rates, and survey costs. We will also examine unweighted demographic distributions across experimental groups.

Economic or Social Exchange: Examining the Effects of Different Incentive Amounts in a Panel Study.

Shelley Boulianne, Grant MacEwan University (sjboulianne@yahoo.com)

Web surveys are becoming quite popular, but researchers are still trying to determine effective methods to improve response rates to web surveys. Compared to mail and telephone surveys, there are relatively few studies of incentives using web surveys.

This study investigates the degree to which the economic exchange theory versus the social exchange theory can explain web survey participation. This study presents the results of an experiment that varied the disbursement of a \$20 incentive in a two-wave panel study (January, March 2010, n=350). In one condition, participants are given a \$10 pre-paid incentive during wave 1, which is followed by a \$10 post-payment after wave 2. In another condition, participants are given a \$5 incentive during wave 1, which is followed by a \$15 post-payment after wave 2. If the economic exchange theory explains survey participation, I would expect to find that response rates at wave 1 are higher for the \$10 incentive group than the \$5 incentive group. Furthermore, I would expect to find that the \$5+15 group would participate at a higher rate than the \$10+10 group in wave 2, because the balance of the reward is greater (\$15 versus \$10). If the social exchange theory explains survey participation, I would expect to find no differences between the two groups in terms of their response rates to both waves of the survey. Furthermore, using wave 1 survey responses about trust, volunteering, and civic ethic, I examine whether the different incentives recruited different types of people during wave 1 and whether these survey responses could be used to predict participation during wave 2.

The findings from this study will help determine effective ways to disburse incentives in a panel design, determine the optimal size of incentives, and understand the motivation for participating in web surveys.

How Much Cooperation is that Incentive Providing? Examining the Benefits of a Single Incentive Across Multiple Surveys.

Alicia M. Frasier, NORC at the University of Chicago (frasier-alicia@norc.org); Heather M. Morrison, NORC at the University of Chicago (morrison-heather@norc.org); Benjamin J. Skalland, NORC at the University of Chicago (skalland-benjamin@norc.org)

Monetary incentives are an important tool for survey researchers attempting to arrest declining response rates, particularly for Random Digit Dial (RDD) surveys where response rates have fallen steadily over the past decade. However, any incentive-driven boost in response rates must be weighed against potential increases in survey costs and bias. As a result, researchers strive to maximize the efficiency of the incentive tool – to obtain as much benefit from each incentive offer as possible.

For this paper, we will explore how much cooperation a single incentive offer can yield. We examine the interaction between incentives and respondent cooperation on the suite of surveys that use the National Immunization Survey (NIS) sampling frame. These surveys include, the NIS Child Survey (NIS-Child), the NIS Teen Survey (NIS-Teen), and the current survey being conducted through the State and Local Area Integrated Telephone Survey (SLAITS) mechanism of the National Center for Health Statistics (NCHS), the 2009-10 National Survey of Children with Special Health Care Needs (NS-CSHCN) sponsored by the Maternal and child Health Bureau (MCHB). Sample lines that are “flagged” for all three surveys proceed in the following order: NIS-Child, NIS-Teen, NS-CSHCN. The NIS-Child, NIS-Teen, and NS-CSHCN all have refusal-based incentive models that can be invoked at any time during the continuum of questionnaires. Following the NIS-Child and NIS-Teen interviews presents the NS-CSHCN with a unique opportunity to capitalize on the preceding incentive efforts. Conversely, offering an incentive in the preceding surveys may lead respondents to expect a further incentive to complete the NS-CSHCN.

We will examine the proportion of respondents who require a further incentive relative to those who do not, factoring selected respondent demographic information into our analyses. We look to shed light on when costs savings could be achieved by maximizing the benefit of the initial incentive offers.

Nonresponse Bias and the Impact of Incentives in a Survey of Physician Members of a Professional Organization

Nicole R Buttermore, American College of Physicians (nbuttermore@acponline.org)

We present data gathered from two randomly-selected samples of 2,000 physician members of a large medical-specialty organization. Sampled physicians were asked to complete one of two surveys that differed in length and content. Participants were first sent links to a web-based version of the survey. While the fields remained open, participants without an email address on file and nonresponders were mailed a paper survey a maximum of four times and were sent links to the web-based version of the same survey a maximum of three additional times. The second paper survey mailing contained a \$2 bill, and the final email and paper surveys each included an offer for a \$10 gift card from an outside vendor. The final response rates to the surveys were 52.3% and 51.2%. Previous research suggests that there are modest amounts of response bias in surveys of physicians (McFarlane, Omsted, Murphy, & Hill, 2007). To test this idea using our samples, we compare responders to nonresponders and internet to paper survey responders using demographic variables available on the sampling frame, such as sex, age, subspecialty, geographic region, primary professional activity, and primary employer. Previous research has also demonstrated that pre-paid monetary incentives are associated with higher response rates in samples of physicians (Kellerman & Herold, 2001). We examine the impact of the pre- and postpaid incentives on response rates to our surveys.

The Effectiveness of Incentives Used in the Second Phase of a Two-phase Survey.

Charlotte L Tubman, Westat (charlottetubman@westat.com); Douglas Williams, Westat (douglaswilliams@westat.com)

With the rise of survey nonresponse, researchers have increasingly included the use of respondent incentives in their data collection protocols as a means to increase response rates. Existing studies have examined a variety of approaches including prepaid versus promised incentives, cash versus non-cash incentives, and varying incentive amounts. The 2009 National Household Education Survey is a large pilot test of 11,800 addresses using a mail mode with telephone followup. Due to the relatively low eligibility rate of households for a topical questionnaire, a two-phase approach is used in which households are screened by mail with possible telephone follow-up (phase 1) and eligible household members are sampled for a topical questionnaire from the completed screeners (phase 2). All households were sent a prepaid cash incentive of \$2 in the initial phase 1 mailing.

This paper examines two experiments concerning the use of incentives in the second phase of this two-phase survey. (1) When the screener was completed by mail and a household member was sampled for the second phase, cases were randomly assigned to receive no cash, \$5, or \$15 with the initial mailing of the topical survey. (2) When households completed the screener by telephone during nonresponse follow-up, the interviewer continued with the topical interview by telephone if a household member was sampled. These cases were randomly assigned to no promised incentive or a promised incentive of \$5 cash upon completion of the topical interview.

This paper will discuss the observed effects of the use and varying amounts of incentives in the second phase survey. As many studies seek alternatives to RDD and may consider two-phase collections when eligibility is low, this research may help to inform incentive procedures.

Methodological Advances in Address Based Sampling

Identifying and Surveying Cell Phone-Only Homes within an Address-Based Sampling Design.
Michael W Link, The Nielsen Company (Michael.Link@Nielsen.com); Jennie Lai, The Nielsen Company (Jennie.Lai@Nielsen.com)

The ability to effectively and efficiently identify and survey cell phone-only (CPO) households has become critical for most studies to ensure the representativeness of the final estimates. To date, however, CPO homes have in large measure been sampled only from known cell phone banks. With the advent of address-based sampling (ABS), -- that is, sampling of addresses from a database with near universal coverage of residential homes -- researchers have an alternative means of identifying and surveying these homes. Using a large ABS mixed-mode study, we explore three key questions. First, how effective are ABS sampling techniques and data collection methods in identifying CPO homes? Second, how does home telephone access differ across various components of the ABS sample, in particular “matched” addresses (those with an identifiable telephone number at the sampling stage) and “unmatched” addresses (those with no identifiable telephone number)? And, third, using a multi-topic questionnaire, how do CPO homes differ from other types of homes across a range of issues (such as consumer behavior, media consumption, and technology use) before and after controls are added for common correlates such as sex, race, age, and education? The data come from a study of 150,000 Nielsen Radio Diary respondents conducted in spring 2009. The findings further both our understanding of the uniqueness (and similarities) of CPO homes versus other types of homes as well as the utility of ABS as a means of sampling CPO households.

Maximizing Response in a Two-Phase Survey with Mail as the Primary Mode

Jill M. Montaquila, Westat (jillmontaquila@westat.com); J. Michael Brick, Westat (mikebrick@westat.com); Mary C. Hagedorn, Westat (MaryHagedorn@westat.com); Douglas Williams, Westat (DouglasWilliams@westat.com)

With recent declines in response rates and challenges in ensuring adequate coverage, survey practitioners are increasingly exploring alternatives to random digit dial (RDD) sampling with computer assisted telephone interviewing (CATI) administration. One such alternative is address-based sampling, using residential address lists based on the U.S. Postal Service Delivery Sequence Files. In the Fall of 2009, we conducted a pilot study to evaluate address-based sampling, with mail as the primary mode of collection, for a periodic survey that had previously been conducted using the RDD/CATI methodology. This study requires two phases of data collection: A Screener to determine a household’s eligibility, followed by a Topical survey administered in eligible households.

The Pilot Study included several embedded experiments with the goal of identifying the most effective approaches for eliciting cooperation. Although mail was the primary mode, various mailing conditions (first class mail for early mailings, and FedEx and Priority Mail for later mailings) were tested, and telephone follow-up was used for a subsample of nonrespondents for whom telephone numbers could be obtained. Additionally, an abbreviated Screener was sent to final Screener nonrespondents, in an effort to obtain information on a very limited set of items.

In this paper, we present the results of the embedded experiments and discuss the implications for an approach that maximizes response in a two-phase survey setting. We also reflect on issues requiring special consideration including the timing of the contacts.

Maximizing a Stratified ABS Sampling Frame for Nationwide Mail Recruitment of a Probability-based Online Panel

Charles A. DiSogra, Knowledge Networks, Inc. (cdisogra@knowledgenetworks.com)

Until recently, the probability-based online KnowledgePanel® run by Knowledge Networks (KN) used exclusively a random-digit dial (RDD) methodology for panel recruiting. RDD-based recruitment has been rapidly losing its technical and cost-effective edge due to the increasing efforts required to achieve acceptable response and coverage rates. More importantly, the shrinking coverage of RDD frames due to the increase of cell-phone-only households and the related disproportionate loss of sample coverage for young adults and Hispanics has rendered RDD recruitment to becoming increasingly inadequate for achieving representative national samples.

An address-based sample (ABS) frame using mail recruitment has emerged as a solution to coverage and minimizing response bias. In 2008, KN conducted a pilot study to recruit panel members via mail using the USPS Delivery Sequence File for its ABS frame. Two large-sample waves were implemented in 2009, supplementing the RDD-sourced sample. Each ABS wave has incorporated refinements to the mailing strategy based on experiments performed in the pilot in prior large-sample waves. These experiments altered materials, incentives and compressed the length of the field period without loss in response rate or sample representativeness. To enhance diversity, the second wave in 2009 included the targeting of high-density minority communities using a stratified sample. Also implemented was a plan to increase the recruitment of households that do not have Internet access. (KN supplies such households with a laptop computer and Internet access so that they, too, can participate in online surveys.) This presentation reviews data demonstrating the evolving effectiveness of KN's mail-based recruitment strategy as evidenced by successive increases in yield and response rate; in improved sample representativeness; in targeting minority areas; and in the recruitment of non-Internet household using a targeting "message" strategy among non-responders.

Enhancing the Computerized Delivery Sequence File for Survey Sampling Applications.

Mansour Fahimi, Marketing Systems Group (mfahimi@m-s-g.com)

Increasingly, survey researchers are considering address-based sampling (ABS) methodologies to reach the general public when proper representation of the target population is a nonnegotiable feature of the study design. Essentially, there are three main factors for this emergence: (1) evolving coverage problems associated with telephone-based methods; (2) eroding rates of response to single modes of contact, along with the increasing costs of remedial measures to counter nonresponse; and (3) recent improvements in the databases of household addresses available to researchers. Virtually, all ABS surveys rely on the Computerized Delivery Sequence File (CDSF) of the USPS as the sampling frame. However, this database is primarily designed for delivery purposes and not sampling applications. As such, CDSF is in need of specific enhancements to improve its sampling and analytical capabilities. This presentation reviews the main reasons for the growing interest in ABS and provides a brief overview of the possible enhancements for the CDSF. While reducing undercoverage, such enhancements enable researchers to develop more efficient sample designs as well as broaden their analytical possibilities through an expanded set of available covariates for hypothesis testing and advanced analytics.

Methodological Briefs: Improving Response Rates

Contingent Incentives Effects on Survey Response

E Nicole Bensky, The Nielsen Company (nicole.bensky@nielsen.com); Anh Thu Burks, The Nielsen Company (anh.thu.burks@nielsen.com); Chuck Shuttles, The Nielsen Company (chuck.shuttles@nielsen.com); Michael Link, The Nielsen Company (michael.link@nielsen.com)

The current incentive structure for many surveys, regardless of amount, involves the use of non-contingent incentives; that is, an incentive that is sent prior to participation. Several previous studies conducted within The Nielsen Company have shown that the use of non-contingent incentives can be an effective approach for improving participation. However, this approach has also been recognized as being very inefficient in that a large percentage of respondents never return the survey even though they were sent an incentive. By moving to contingent incentives (combined with an initial modest non-contingent incentive) the focus is on rewarding respondents that perform the requested task. Contingent incentives also provide an opportunity to offer responding homes a larger incentive than they would have received under the current up-front incentive plan of smaller amounts to all survey recipients. The use of these larger contingent incentives is hypothesized to have a positive impact on participation rates and provide many cost efficiencies. In February 2010 a test was conducted involving the use of contingent incentives to help improve the responses rate within the TV Ratings Diary. For this test, approximately 32,500 households were mailed dairies using differential incentives with higher incentive amount for those households identified as having key demographics (such as younger ages of householder, Black race, or Hispanic ethnicity). Initial non-contingent incentives (\$1 or \$2) were sent to households along with the diary packet that also contained mention of a higher incentive once the dairies were completed and returned. Contingent incentives also depended on household demographics (amounts either \$20 or \$50) and were mailed when at least one non-blank diary had been received by the household. In the analysis, we examine (1) the effectiveness of the use of contingent incentives with varying incentive amounts on cooperation and (2) the costs impacts of larger contingent incentives.

Getting a Physician's Attention: Results of an Upfront Incentive Mail Survey Experiment

Jordon Peugh, Knowledge Networks (jpeugh@knowledgenetworks.com); Sandra Applebaum, Harris Interactive (sapplebaum@harrisinteractive.com); Brenda Sirovich, Outcomes Group, VA Medical Center (Brenda.Sirovich@Dartmouth.edu)

Physician survey research has increasingly been plagued by poor response rates. Previous work has shown that upfront incentives - provided at the time of an invitation to complete a survey – boost response rates, but it is not known whether cash outperforms checks and what dollar amount is sufficient. We conducted an experiment to examine the impact of four different upfront incentives on response to a physician mail survey.

In the spring of 2009, we mailed a nine-page questionnaire on practice intensity to 1,000 primary care physicians randomly selected from the Physician Masterfile of the AMA. We randomly assigned physicians to receive one of four incentives: \$20 cash; \$50 cash; \$50 check; \$100 check.

The primary outcome measure was response rate at two months after the initial mailing. We also examined response rates over time up until four months. At two weeks after the initial mailing, the \$50 cash incentive achieved the highest response rate at 40%, higher than both the 32% achieved with \$20 cash ($p=0.05$) and 29% with the \$50 check ($p=0.01$), and comparable to 36% with the \$100 check ($p=0.3$). At two months, the \$50 cash incentive again had the highest response rate (56%), higher than the \$20 cash option (48%, $p=0.03$) and the \$50 check (50%, $p=0.07$) and comparable to the \$100 check (55%, $p=0.5$). A similar pattern persisted through four months, by which time each of the upfront incentive options was associated with a response rate of 50% or greater.

We achieved a 50% mail survey response rate among primary care physicians with any of four upfront monetary incentives. Individual researchers must decide whether the marginally higher response rates achieved with higher incentive amounts are worth the additional cost. The choice of cash versus check should depend on expected response and check cashing rates and investigator comfort.

Differential Mailing Methodologies on Consent Rates: Testing USPS vs. UPS for Pre-Recruitment Mailing.

Yelena Pens, Arbitron (yelena.pens@arbitron.com); Adam Gluck, Arbitron (adam.gluck@arbitron.com)

Arbitron Inc. has developed the Portable People Meter (PPM™), a new technology for media and marketing research. A panel-based methodology is used to collect PPM™ related data. Panel recruitment procedures include a multi-step, multi-mode process in order to optimize cooperation. As a first step prospective panelists are asked to participate in a short survey about their media consumption habits.

Those who agree are mailed a promised incentive, as well as recruitment materials inviting the members of the household to participate in the PPM panel.

In the spring of 2009, Arbitron conducted a test to determine if sending the promised incentive and recruitment materials via USPS first class mail would impact our panel recruitment agree rates, as compared to sending them via our current methodology of overnight UPS. In this test, several hundred household's that agreed to complete our short initial survey were sent a recruitment mailing via an overnight UPS mailing, while another several hundred were sent the recruitment mailing via USPS.

In this paper, the differences between the recruitment agree rate of the overnight UPS mailing group and USPS mailing group are analyzed to determine whether a delay in the receipt of the promised incentive and recruitment materials affects prospective panelists' willingness to agree to be a part of the panel.

The Use of Monetary and Nonmonetary Incentives to Increase Response Rates in a Mail Survey of African-American Mothers

John Stevenson, University of Wisconsin Survey Center (stevenson@ssc.wisc.edu); Jennifer Dykema, University of Wisconsin Survey Center (dykema@ssc.wisc.edu); Chad Kniss, University of Wisconsin Survey Center (ckniss@ssc.wisc.edu)

Infant mortality rates in the United States are alarmingly high for a developed nation, particularly among African Americans. In Wisconsin, African American infant mortality rates are among the worst in the US. The Pregnancy Risk Assessment Monitoring System (PRAMS), sponsored by CDC and state health departments, is designed to collect high-quality data on pregnancy and infant health not found in other data sources. However, response rates among African American mothers have been consistently lower than for non-Hispanic white mothers and in 2009, an intervention was implemented to increase participation in which African American mothers were randomly assigned to one of the following groups:

- Group 1: control group that received no incentive
- Group 2: received a coupon for diapers (cash value of \$6)
- Group 3: received a pre-paid \$5 cash incentive

Incentives were included with the Wisconsin PRAMS survey which is mailed to respondents. The mail portion of the survey has five waves including an advance letter, a reminder letter, and up to three copies of the survey. If study participants have not completed the survey after two months they are included in a phone follow-up phase to increase response.

The analysis has three parts: the effects of the experimental treatments on unit and item nonresponse; the effects of incentives on survey reports to assess nonresponse bias, exploring differences among groups in survey responses (i.e. basic demographics) and differences between responders and

nonresponders using administrative data; and an analysis of the cost variation among the groups. Preliminary results indicate the pre-paid \$5 cash incentive significantly increased response rates: 59% in that group compared to 49% in the coupon group and 42% in the group that did not receive an incentive. This analysis adds to research on the effects of incentives in mail surveys, particularly among minority and low-income groups.

Differential Incentives Within Households: A Long-Term View.

Courtney N Mooney, Arbitron (courtney.mooney@arbitron.com); Ryan McKinney, Arbitron (ryan.mckinney@arbitron.com)

For its Portable People Meter (PPM) panel, Arbitron has historically provided all panelists within a household the same monetary incentive. Previous research has demonstrated that higher monetary amounts are necessary to encourage young adults to participate in the panel, but providing all household members who live with a young adult the higher monetary incentive is neither cost-effective nor efficient. If incentive dollars could be targeted to the young adult, without negatively impacting the participation of other household members, this would result in substantial cost savings. In 2008, Arbitron conducted a brief test to evaluate the feasibility of differential incentives in households with the presence of at least one 18-24 year old. In the test households, the 18-24 year old received a higher monetary incentive than other household members; whereas in the control households, all household members received the same incentive as the 18-24 year old. The results of this test indicated that the differential incentive structure could be applied in specific household types without decreasing compliance and Arbitron implemented the practice in early 2009. This presentation will examine the long-term effectiveness of the initiative. Specifically, compliance rates and panelist tenure will be analyzed to determine whether there is variation in the acceptability of differential incentives by household size and composition.

Do Response Rates Matter? Results of a Controlled Experiment

Ellen Denzen, National Marrow Donor Program (edenzen@nmdp.org); Todd Rockwood, University of Minnesota (rockw001@umn.edu); Nancy Omondi, National Marrow Donor Program (nomondi@nmdp.org); David Radosevich, University of Minnesota (davidmr@umn.edu); Elizabeth Murphy, National Marrow Donor Program (emurphy@nmdp.org)

Historically response rates have been used as one of the primary indicators of the quality of survey data. Declining rates of response across survey modes have made meeting this standard difficult; the issue of nonresponse bias has been increasingly cited as a potential risk to data integrity. This study assessed the effects of Tailored Design Method (TDM) survey implementation on response rate and nonresponse bias. The National Marrow Donor Program conducted the Patient Satisfaction Survey (PSS), a cross sectional survey of patients that searched the Be The Match RegistrySM from July – August, 2009 (N=800). Subjects were randomly assigned to either the control group which received only a single mailing of the survey or the experimental group that received: 1) pre-notice letter followed by mail survey with \$2 prepaid incentive, 2) follow-up survey for nonrespondents, and finally 3) telephone interview for persistent nonrespondents. The control group had a response rate of 12.8% compared to a response rate of 50.5% in the experimental group ($p < .05$). Over the past five years, the PSS has had an average response rate of 11%. Furthermore, we were able to compare basic demographic differences between the control and experimental groups. The TDM was successful in bringing in more respondents over the age of 50 (OR 0.52) and those who had an 'other diagnosis' (e.g. hematological diseases, immunological deficiencies and genetic disorders treatable by blood and marrow transplant such as sickle cell disease or Fanconi anemia; OR 0.56). While the TDM did produce a much higher response rate, comparison of the data between the control and experimental groups showed no significant differences for substantive variables.

Public Opinion & the Health Care Debate

Framing the News about Health: Effects of Controllability Attributions in the Media on Public Opinion and Health Policy Preferences

Jocelyn Landau, University of Pennsylvania (jocelyn.landau@uphs.upenn.edu)

News media framing plays an important role in shaping public opinion about modern developments in health and science. Greater attention to controllable risk factors for diseases can increase the public's awareness about prevention methods and shift policy attention to non-medical strategies to improve population health. At the same time, advances in molecular genetics have brought greater attention to the role of genetics in assessing disease risk. Whereas behavioral and genetic factors are both influential in disease prevention, the former is a controllable risk factor and the latter is uncontrollable. To date, little is known about how the public responds to media framing that emphasizes the controllable versus uncontrollable nature of common diseases.

This study employs an Internet-based survey experiment to test the differential effects of news framing that highlights controllable (behavioral) risk factors for disease versus uncontrollable (genetic) risk factors. This research tests whether controllability framing in news media coverage about heart disease influences Americans' opinions about current health topics and related health policies. Participants were from a nationally representative sample of American adults (N = 1,602). Findings reveal that framing heart disease risk as either controllable (behavioral) or uncontrollable (genetic) influences peoples' opinions, causal attributions for disease, and health policy preferences. The study also shows that exposure to the controllable message frame led to a divergence between Conservatives' and Liberals' policy opinions, relative to their opinions after exposure to the uncontrollable message frame. The results suggest that increasing public awareness about the preventable nature of common diseases may have a polarizing effect across ideological lines, resulting in substantially lower policy support among Conservatives.

Polling on a Fuzzy Issue Like the 2009 Health Care Reform Debate: A Suggestion Of “Do’s and Don’ts” From the Trenches.

Clifford Alexander Young, Ipsos Public Affairs (clifford.young@ipsos.com)

Polling on the health care reform debate has been challenging to say the least. On any given day, different people were able to cite different polls and come to very different conclusions.

In our opinion, the central problem has been our underlying assumption that polling on non-electoral policy issues involving the legislative process is analogous to polling on presidential elections. Why?

In presidential elections, our job as pollsters is made easy, with ballot questions being basically fixed after the primaries. In contrast, issues like health care reform are quite fuzzy as no bill typically exists at the beginning of the process. Put another way, we have no “true value” to measure against.

What does polling on American health care reform teach us about polling on non-electoral policy issues involving the legislative process?

To answer the above question, we analyze six Ipsos-McClatchy polls on health care and two separate wording experiments, which are all nationally representative RDD surveys of the American population.

Our central conclusion is that no one single ballot question is superior. Instead, fuzzy issues like health care require a multiple indicators approach.

Specifically, generic questions, like “general support for health care proposals”, are very vague with variable meanings depending on the person’s given perspective. Indeed, a simple follow up question shows that about a quarter of people “opposed” health care reform because “it did not go far enough”. For these very reasons, generic ballot questions are highly susceptible to question wording (government-run, government administered Obama’s Plan, etc.).

In contrast, specific ballot questions, like those on the “public option”, are hypothetical at best and, as such, should be used more for sensitivity analysis than for predictive purposes.

We also include a list of “do’s and don’ts” to help polling on future non-electoral reform debates.

Support for Health Care Reform: It All Depends on How You Ask the Question.

Trevor Thompson, The Associated Press (ttompson@ap.org); Jon Krosnick, Stanford University (krosnick@stanford.edu); Dennis Junius, The Associated Press (djunius@ap.org); Josh Pasek, Stanford University (josh@joshpasek.com)

Health care reform has been the hottest topic in Washington, and has also been a hot topic for public polls this year. But on a topic that is so complex, with so many different plans and options being discussed, polls have produced differing results on whether the public favors or opposes health care reform. We present results from an extensive Associated Press polling project, conducted by Stanford University with the Robert Wood Johnson Foundation, to illustrate the significant difference that question wording can make in the results of opinion surveys about health care reform. We also collected a vast array of questions and results from various other survey organizations as part of our analysis. The AP surveys, including a telephone survey conducted by GfK Roper and an online survey conducted by Knowledge Networks, contained several question wording experiments and different question constructions measuring similar concepts. Overall, we find that the public expressed much higher support for various proposals for health care reform when the questions were worded more generally, while support was much lower when more specific information was presented to the respondents about the pros and cons of each proposal.

Who Supports Health Care Reform? Explaining the Determinants of Support for Various Health Care Reforms

Josh Pasek, Stanford University (josh@joshpasek.com); Trevor Thompson, Associated Press (ttompson@ap.org); Jon A. Krosnick, Stanford University (krosnick@stanford.edu)

Health care reform has been a dominant issue in American politics throughout 2009. The health care debate is exceedingly complex, with a large number of proposals that individuals could support or oppose. Using data from a survey by the Associated Press and Stanford University, with support from the Robert Wood Johnson Foundation, we explore the determinants of support for health care reform in general, support for the plans being discussed in Congress, support for the public option, and preferences on 13 other proposals and goals for health care reform.

In particular, we examine (1) whether the same set individuals support all of the plans, (2) the extent to which support for various plans is simply a function of partisanship and ideology, (3) whether support for the plans varies based on what individuals expect from Congressional reform, and (4) whether perceptions of and experiences with the health care system are related to desire for reform. We identify a

series of significant differences between various predictors of health care reform indicating that Americans do seem to evaluate each proposal in its own light. We explore each of these differences and compare the specific and more general measures to show how Americans seem to be considering health care reform.

The Effectiveness of Consumer and Provider Education and Outreach About Health IT: A Content Analysis.

Stephanie Rizk, RTI International (srizk@rti.org); Nicole Knops, RTI International (nknops@rti.org)

The focus of this research is to measure the effectiveness of outreach related to public understanding and opinions regarding electronic health information exchange (HIE) within and across various states.

The Health Information Security and Privacy Collaboration (HISPC), established in January 2008 by RTI International through a contract with the U.S. Department of Health and Human Services (HHS), assembled 42 states and territories to address the privacy and security challenges presented by electronic HIE through multi-state collaborative workgroups. Two of these multi-state collaboratives focused on developing tools and strategies to educate stakeholders about HIE; one to engage consumers and another to develop a toolkit to educate providers. Between April and July 2009, states participating in the HISPC were challenged to test the effectiveness of these consumer and provider education materials by adapting and disseminating the materials within their communities.

To measure effectiveness of this outreach across the various states, we will perform a content analysis of final reports from each state using qualitative analysis. The information will be separated into two major categories: accomplishments and limitations. Within these categories data will be coded to indicate differing outreach methods, number of organizations/individuals involved in each separate initiative, and number of commitments to officially disseminate the materials. Within the coded data we will look both within and across the states to determine how widely adapted and disseminated the materials were.

Initial reviews of the material indicate that most participating states achieved significant success in adapting and disseminating the tools originally created by the multi-state workgroups for both provider and consumer education. However, some states reported limitations that may pose important lessons and future directions in creating materials intended to shift public opinion by using a multi-component workgroup methodology.

AAPOR Poster Session 2

Friday, May 14, 2010

3:15 p.m. - 4:15 p.m.

Poster Session - 2

Evaluating Response Rate and Response Bias Changes After Implementing a Multi-Mode Survey and an Electronic Survey Record System

Rebecca Lien, Professional Data Analysts, Inc. (blien@pdastats.com); Lija Greenseid, Professional Data Analysts, Inc. (lija@pdastats.com); Barbara A Schillo, ClearWay Minnesota (bschillo@clearwaymn.org); Raymond Boyle, ClearWay Minnesota (rboyle@clearwaymn.org); Theresa Leonard, ClearWay Minnesota (tleonard@clearwaymn.org); Michael G. Luxenberg, Professional Data Analysts, Inc. (michael@pdastats.com)

This study examines response rates and response bias of adult tobacco cessation program participants under different modes of survey administration. Follow-up surveys were conducted 7 months after enrollment. We initially conducted surveys using a single mode (telephone) with interviewers using paper call sheets for tracking. We recently introduced a multi-mode survey using telephone with mail follow-up to non-responders – including a \$2 cash incentive. Two months after these changes, we implemented an electronic call tracking tool for interviewers. This poster compares response rates and response bias of three cohorts: 1) single mode with paper tracking, 2) multi-mode with paper tracking, and 3) multi-mode with electronic tracking.

Methods: We collected participant intake and follow-up data for this observational study. Cohort-1 participants enrolled between March and September 2007 (n=886); Cohort-2 participants between April and May 2008 (n=150); and Cohort-3 participants between June 2008 and May 2009 (n=1210). Chi-squared and t-tests were performed.

Results: The response rate improved from Cohort-1 (68.8%) to Cohort-3 (76.4%). Cohort-2 (73.3%) had a non-significant increase from Cohort-1. Response bias analysis of Cohort-1 intake variables found that responders were more likely to be white and older than non-responders. Cohort-3 responders were also more likely to be white and older than non-responders; in addition, Cohort-3 responders were more highly educated. The main program outcome – quit rate – for Cohort-1 (31.6%) was higher than for Cohort-3 (25.1%).

Discussion: Response rates improved after implementing new survey modes. We believe mail surveys provided both an alternate response mode and extra incentive. In addition, electronic tracking provides greater efficiency and accuracy when managing survey calls.

Response bias among intake variables was not improved. However, Cohort-3 while having a higher response rate had a worse outcome than Cohort-1. Other factors not controlled for in this observational study may have affected response rates and response bias.

Social Networking Websites: Static Online Ghettos or Evolving Melting Pots?

Katy Schwalbe, Zogby International (katy@zogby.com); Zeljka Buturovic-Bradaric, Zogby International (zeljka@zogby.com); Grace Ren, Zogby International (grace@zogby.com)

Recent qualitative ethnographic research, conducted primarily in 2005 and 2006, raised the possibility that racial and income divides exist among users of social networking sites MySpace and Facebook. Much of this work, and most of the mainstream media reporting, focused explicitly on racial distinctions: African American youth appeared to be more likely to be MySpace users, while nonhispanic white youth gravitated towards Facebook.

However, it is unclear whether this pattern of use presents a permanent cyber-space segregation, or a difference due to a time lag in adopting new technologies among different demographic groups, which could gradually disappear. Results of six surveys conducted by Zogby International over the past year regarding social networking preferences, usage, opinions, and habits expand on both hypotheses. Our data corroborate the existence of a preference for MySpace over Facebook among African American youth with a corresponding preference for Facebook over MySpace use among nonhispanic white youth. However, income displays a more nuanced pattern: while low-income youth are less likely than youth overall to use Facebook, among low-income youth who do use social networking sites, more use Facebook than MySpace. Our existing surveys do not include a sample size large enough to specifically analyze social networking use among low-income minority youth; a Zogby survey in the field contains an oversample of this group for later analysis and comparison. Finally, a time-series analysis of responses to social networking questions over the previous year suggests a small but statistically significant overall migration away from MySpace and towards Facebook among youth of all races and income groups, suggesting that the “ghetto-ization” of social networking sites will become less of a factor as most racial and income groups migrate to the same popular alternative.

Sampling Strategies for Face-To-Face Interviewing in a Racially Diverse Metropolitan Area
Kelly Ann Marzano, Survey Research Laboratory, University of Illinois - Chicago
(kellym@srl.uic.edu); Karen Foote Retzer, Survey Research Laboratory (karenr@srl.uic.edu)

In this presentation we discuss sampling challenges for a study involving phone recruitment and face-to-face interviews in a major metropolitan area. The study sought to explore how cultural orientation affects survey response behaviors in four racial/ethnic groups in the Chicago area, including Whites, African Americans, and respondents of Mexican and of Korean descent. For the first three groups, the initial sampling plan used listed sample in the city of Chicago from tracts where a high proportion of the population was known to be eligible. For the Korean sample, the initial plan relied on listed sample using the last name Kim. Kim is the most popular Korean surname and previous research indicated this was an effective method of recruitment.

Our experiences quickly demonstrated we would need to modify our sampling strategies. The initial strategy for recruiting the White, African-American, and Mexican sample proved inefficient. We instead experimented with sampling based on public transportation routes. The initial challenge regarding the Korean sample was that there were not enough Kim or other Korean surnames in Chicago or even Cook County as a whole. As a result, we explored other methods of sampling such as snowball sampling, posting in on-line forums, newspapers, and community centers, and off site interviewing in Korean communities.

We will present results on the overall cooperation by sampling strategy and based on those who set face-to-face appointments by each of the racial/ethnic groups that we recruited. For the Mexican and Korean sample we will make comparisons between the English-speaking and Spanish- or Korean- speaking samples.

When a Single Measure is Sufficient: Optimizing Survey Efficiency in Concept Evaluation Research.

Jennifer G. Boldry, National Analysts Worldwide (jboldry@nationalanalysts.com)

In order to minimize risks to data quality associated with respondent burden and fatigue, the market research industry has become increasingly concerned about balancing survey scope and length. Inspired by these concerns, National Analysts Worldwide has applied meta-analytic techniques to provide empirical guidance in improving the efficiency of survey construction in concept evaluation research. Meta-analysis is a robust method of evaluating the efficacy of study variables through aggregating and analyzing the magnitude of effects across a number of independent surveys. In the present research, we meta-analytically examine relationships among metrics that are commonly included in concept testing survey research (i.e. ratings of how “compelling”, “credible”, and “unique” the concept is) in order to understand whether these concepts (a) measure independent constructs and (b) predict the likelihood of a customer to act (i.e. seek additional information or purchase a product/service). Results based on the data from 10 independent surveys using a variety of sampling techniques across three industries (Healthcare, Information Technology, and Transportation) suggest that (1) ratings of “compelling,” “credible”, and “unique” are correlated with one another, and thus that (2) use of all three metrics as predictors of “likely to take action” nets little incremental gain in predictive validity over the most correlated metric as a single predictor (i.e., compelling). Recommendations for optimizing survey efficiency in concept evaluation research are discussed.

Considering Complementary Evaluation Methods to Augment Cognitive Techniques During Pre-testing of Establishment Surveys

Kenneth M Pick, U.S. Census Bureau (kenneth.pick@census.gov); Alfred D Tuttle, U.S. Census Bureau (alfred.d.tuttle@census.gov); Diane K Willimack, U.S. Census Bureau (diane.k.willimack@census.gov)

Pre-testing of establishment surveys at the U.S. Census Bureau predominantly uses a mix of cognitive interview techniques, as they are effective in uncovering problems related to survey concepts, comprehension of survey items, and retrieval of information from records. However, some problems uncovered cannot be solved through cognitive testing alone and require the use of other complementary pre-testing methodologies.

This research reviews findings from the use of cognitive interviews in pre-testing establishment surveys conducted over the past five years at the Census Bureau. Through in-depth interviews with Census Bureau researchers, we identify potential problems discovered in these establishment instruments that could not be thoroughly evaluated using cognitive interviews alone. We then propose both qualitative and quantitative methodologies that could complement cognitive interviews to more fully understand the impact of recommendations for changes to survey instruments.

The results indicate that cognitive interviews are effective in pre-testing establishment instruments. However, when complementary pre-testing methodologies are used in concert with cognitive interviews, they may offer the ability to corroborate cognitive interview findings or suggest the need for further questionnaire development and pre-testing.

Explaining Selective Exposure: The Impact of Opinion-Perception Incongruity on Media Attention in the 2004 Presidential Campaign

Lindsay H. Hoffman, University of Delaware (lindsayh@udel.edu); Rebecca Riley, University of Delaware (reriley@udel.edu)

Much research has found that perceptions of public opinion have an impact on interpersonal behaviors, media use, and political expression (Hardy & Jamieson, 2005; Hayes, 2007; Knobloch-Westerwick & Meng, 2009; Mutz & Martin, 2001; Noelle-Nuemann, 1993; Scheufele & Eveland, 2001). At the same time, cognitive dissonance theory and selective exposure research have revealed what happens when people are confronted with information counter to their beliefs (Festinger & Carlsmith, 1959; Jang, 2009; Mutz, 2002). This study focuses on the intersection of these research areas, examining the incongruity between one's own opinion and his or her perceptions of others opinions, and subsequent effects on media behaviors. We analyzed this phenomenon in the context of the 2004 presidential election, hypothesizing that opinion-perception incongruity would predict media attention.

Using 2004 ANES data, we created an opinion-perception incongruity measure by categorizing responses to intended candidate support and perception of public opinion. Respondents were classified as having either high or low incongruity based on who they supported and who they thought would win, for a total of four categories. Controlling for gender, education level, income, age, ideology, political interest, and political efficacy, we ran a hierarchical regression analysis to determine the impact of opinion-perception incongruity on attention to television. Results suggested that people who supported Kerry and thought he was going to win paid significantly more attention to television news than people who both supported Bush and thought he was going to win and supported Bush and thought he was going to lose. Interestingly, people who supported Kerry and thought he was going to lose displayed no significant differences in attention paid to television. Results are discussed in the context of the 2004 election, and implications for theory and public opinion research are discussed.

Surveying Across Multi-Ethnic Communities in Los Angeles: Lessons Learned.

Karen L Tucker, Battelle (tucker@battelle.org); Hong Li, Battelle (LIH@battelle.org)

Surveys of diverse racial and ethnic groups face distinctive methodological and empirical challenges, including issues related to defining ethnic identity, developing survey questions for multi-ethnic populations, and locating, recruiting, and communicating with the groups of interest. Since 1998, Battelle Centers for Public Health Research and Evaluation and the University of Southern California have partnered to recruit survey and conduct clinical eye examinations on over 19,000 respondents in the greater Los Angeles area. These studies focused on Latino, Non Hispanic White, African American, and Asian American populations, each comprised of diverse individuals who differ in their linguistic, cultural, and socioeconomic backgrounds and English proficiency. We have identified valuable factors to increase validity, efficiency, and participation when surveying across multi-ethnic communities.

- Allow participants to self-identify ethnicity by selecting all Census categories that apply;
- Geocode addresses to identify areas with highest density of target population;
- Conduct key informant interviews and or focus groups to inform project design and community outreach;
- Drive through targeted areas to identify visible signs supporting choice of survey location;
- Provide ongoing community engagement by reaching out to cultural and community centers and leaders, new immigration resources, translation services, faith-based organizations, etc.
- Obtain support of respected community leaders to enhance acceptance of survey within target communities
- Attend community fairs or native tradition-related activities to facilitate information exchange;
- Translate instruments and study materials into appropriate languages, dialects and expression styles;
- Provide advance materials in English and languages of households within targeted areas;
- Hire bilingual staff who speak population-specific native languages from within the community to facilitate recruitment and data collection;
- Maintain flexibility in scheduling appointments remaining cognizant of cultural holidays;

- Offer transportation that is acceptable to participants;
- Provide clinic examination results in English and preferred language of participant.

Rethinking LA: Attitudes and Outlooks After the Riots.

Brianne Gilbert, Loyola Marymount University (brianne.gilbert@lmu.edu); Melissa Woehrstein, Loyola Marymount University (melissa.woehrstein@lmu.edu)

On April 29, 1992, three white and one Latino police officers were acquitted by a majority white jury for the beating of African American motorist Rodney King. Situated at the beginning of the 1990's recession, gang violence in the City of Los Angeles was already at a high, and with the added fuel of these and other racially charged events, the LA Riots broke out soon after the verdict was read. During the week-long Riots, more than 50 people were killed and over 1,000 buildings devastated, resulting in nearly a billion dollars of damage. Many wondered how these Riots would affect future quality of life. Race relations in LA could no longer be ignored.

In observance of each of the 5, 10, and 15 year anniversaries of the LA Riots, the Leavey Center for the Study of Los Angeles (LCSLA) at Loyola Marymount University sponsored cross-sectional phone surveys of Angelenos to study their attitudes toward Los Angeles. The 15 year anniversary survey incorporated responses from 1,651 randomly and ethnically represented residents in the city of Los Angeles (margin of error $\pm 3\%$). Similar to previous surveys, the 15th Anniversary Resident and Interview Surveys provided longitudinal data on six main categories: 1) race/ethnic relations; 2) riot activity; 3) government/community relations; 4) immigration; 5) social class and the American Dream; and 6) demographics.

Results of univariate analyses of the question categories described above and multivariate analyses describing the relationship between demographic shifts, multi-ethnic relationships, riot activity, and the impact of the Riots on Angelenos' quality of life are explored. Although the LA Riots will be remembered as a tragedy, LCSLA is looking beyond the city's diversity as a source for racial tension to provide LA residents with an opportunity to express their attitudes, concerns, and hopes for their city.

Why So Hard to find? Using State-Level Cell-only, Unemployment, and Immigration Rates to Assess Panel Survey Tracking Success.

Julia McQuillan, University of Nebraska-Lincoln (jmcquillan2@unl.edu); David R Johnson, PennState University (drj10@psu.edu); Andrew V. Bedrous, University of Nebraska-Lincoln (abedrou2@unlserve.unl.edu); Stacia Jorgensen, University of Nebraska-Lincoln (shalada2@unlnotes.unl.edu); Ashley J. Frear-Cooper, University of Nebraska-Lincoln (afrear2@unlnotes.unl.edu)

Panel studies are critical for many research projects, particularly those focused on health topics. Tracking and re-contacting participants is essential to the success of longitudinal research. Yet, retaining participants over time has been difficult, even for studies that had complete information on participants at baseline, including social security numbers. Several studies have focused on declining response rates and potential biases from cell-only households for random digit dial (RDD) telephone studies. Less research has focused on contemporary challenges for tracking respondents because of changes in technology (e.g., transitions to cell-only households), changes in the economy (e.g., higher unemployment and related foreclosures), and changes in policies (e.g., immigration crackdowns). Using the National Survey of Fertility Barriers (NSFB), a two-wave panel study of women aged 25 to 45 and their partners first interviewed between 2004 and 2006 and re-interviewed between 2007 and 2009, we use state-level variations in contact with participants along with rates of cell-only households, unemployment, and immigration to seek explanations for challenges to follow-up interviews. We are

currently unable to assess concerns about identity theft, time pressures, relationships ending, or other potential threats to continued participation in this study. Results from bivariate associations from this ongoing study show that recent technological, economic, and political changes in the United States are associated with differential odds of contact and cooperation. As anticipated, states with higher rates of cell-only households, unemployment, and immigration have lower contact rates than states with lower levels of all three measures. Next steps will involve adding individual-level characteristics to the state-level analysis using multilevel models of survey cooperation and completion. We are particularly interested in lower participation in follow-up interviews from Black and Hispanic compared to White participants.

Results of Lab and Field Testing on Questionnaire Development

Patricia Sanchez, U.S. Census (patricia.k.sanchez@census.gov); Travis J Pape, U.S. Census Bureau (travis.pape@census.gov)

The Person Followup (PFU) interviewing phase of the Census Coverage Measurement (CCM) program involves the followup of persons to resolve inconsistent information between data from the CCM and from the census using a paper questionnaire.

The main goal of our tests was to evaluate changes to the introduction, which includes questions to help the interviewer identify a respondent. In order to find a layout that would work in all situations, we went through several phases of testing. Following the 2006 Census Test, we made some revisions to the layout based on interviewer debriefings. We then conducted two rounds of cognitive testing in a controlled lab environment to see if interviewers would be able to follow the layout of the form. These interviewers had an hour of training and conducted three interviews each.

To test how well the introduction worked in complicated cases, a field test was also needed since these conditions are difficult to reproduce in a lab setting with an abbreviated training. We conducted a small field test where volunteers attended two days of training and then conducted about 30 interviews each. Each interviewer was paired with an observer familiar with the operation. These observers recorded any issues seen with the form and procedures as well as conducted respondent debriefings in cases where it was difficult to capture the true situation.

Based on the results of the field test, additional changes were made to the introduction and a second field test was conducted to test these new changes. This paper will describe the methods used in the initial field test, results that were found and questionnaire changes made. We will then review the second field test and the resulting questionnaire changes.

Who Are Latino Cell Phone Only Respondents? A Comparative Look at Demographics, Attitudes, and Behaviors of U.S. Hispanic and Non-Hispanic Cell Phone Only and Landline Respondents
Martin G. Cerda, Encuesta, Inc. (martin.cerda@encuesta.com); Ilgin Basar, Encuesta, Inc. (ilgin.basar@encuesta.com)

Over 20% of all U.S. households now exclusively use cell phones for voice communication (National Center for Health Statistics, "National Health Interview Survey", July-December 2008).

U.S. Hispanics are more likely to "only have a cell phone" compared to white non-Hispanics: 31% vs. 22%; respectively. Furthermore, among those who have both cell and landline phones, English-speaking Hispanics are more likely than white non-Hispanics to say they make most of their calls on their cell phones (Pew Internet & American Life Project, "Cell Phone Use", April 2006).

Given this change in society, it is crucial to include a “cell phone frame” when designing telephone methodology based surveys, in particular among Latinos across all acculturation levels. Yet, who exactly are cell phone only respondents? How do they differ from landline phone owners? Understanding the characteristics of respondents drawn from cell phone and landline sample frames and meeting the “cell phone only” definition may be critical when interpreting research findings on any given topic.

To this end, Encuesta, Inc. conducted a nationally representative study (November 6-27, 2009) with n=358 Hispanic and n=302 non-Hispanic adults using a random probability landline (RDD and listed Hispanic surname) and cell phone (RDD) frame sampling approach. The study was conducted “in language” and “in culture” in order to gain proper representation of all U.S. Hispanics by language usage and acculturation level. A comparative analysis will highlight useful differences between the two sample frames and among “cell phone only” Hispanics vs. non-Hispanics on demographics, cell phone usage, the future of landline phone ownership, and respective views on a variety of topics including corporate responsibility, healthcare, economic conditions, and politics.

Extreme Response Style among U.S. Hispanics vs. Non-Hispanics: An Experiment of Rating Scale Usage and Implications for Survey Design and Analysis.

Martin G. Cerda, Encuesta, Inc. (martin.cerda@encuesta.com); Ilgin Basar, Encuesta, Inc. (ilgin.basar@encuesta.com)

It is well documented that U.S. Hispanics have a tendency to give extreme responses more so than white non-Hispanics. (Gerardo Marin, Raymond J. Gamba, Barbara V. Marin, “Extreme Response Style and Acquiescence among Hispanics”). However, there is no consensus as to how to best measure extreme response style among Latinos and control for its effects in the data analysis and interpretation.

To shed new light, Encuesta, Inc. carried out an experiment as part of a nationally representative study between November 6-27, 2009 with n=358 Hispanic and n=302 non-Hispanic adults using a telephone methodology and a random probability sampling approach. The study was conducted in such a manner as to obtain proper representation of all U.S. Hispanics by language usage and acculturation level.

The interviewing process had two phases, testing and retesting of six experimental “bedrock” questions (i.e. opinions should not shift over a brief period of time).

For the testing phase, questions related to several topics were asked with a mixture of 5-point Likert and 10-point anchored scales. This would help reduce “response set effects” as the use of same response format was avoided (C. Harry Hui, Harry C. Triandis, “The Instability of Response Sets”).

In the retesting phase, half of the respondents were asked the six “bedrock” questions but with a different scale. Specifically, some of the retest questions were converted to 5-point and some to 10-point scales. The other half of the respondents were retested to assess the reliability of the six “bedrock” questions using the exact scales from the testing phase.

The findings will provide evidence of extreme response style and highlight differences and possible solutions when conducting research among Hispanics.

“Testing Recruitment and Retention Strategies for Longitudinal Health Research on Chicago’s South Side”

Kevin V Ulrich, University of Chicago Survey Lab (ulrichkv@uchicago.edu)

The success of a longitudinal study of individual health outcomes is dependent on the ability to effectively sample, recruit and retain members of the population being studied. The challenge is exacerbated when dealing with traditionally marginalized populations or populations who may be hesitant to engage with the organization conducting the research. Phone coverage is low in inner-city neighborhoods and cell phone samples cannot be targeted to specific community areas. As a result, these populations are often under-represented in longitudinal, health-related studies which may lead to disparities in research quality. In order to improve data quality through more representative samples and higher response rates, novel methods must be tested and employed to successfully recruit and retain persons from marginalized populations.

The University of Chicago Survey Lab is conducting a pilot survey to test incentive and mode preferences in two low-income areas on Chicago's South Side. Cases were sampled in order to specifically target persons most difficult to recruit and retain for longitudinal studies: those who live in single room occupancy (SRO) residences and efficiency, low-rent apartment buildings. The outcomes of these cases will be compared with the outcomes of probability cases sampled from address-based samples within the same communities. The study includes both pre- and post- incentives. Cellular phones were also offered to participants as incentives and also as a means to test the device's ability as a recruitment and retention strategy. The study tests the level of response that can be achieved with prompting across multiple modes for interview-administered phone or in-person surveys.

This paper will discuss the findings of the pilot survey in terms of recruitment, retention and representation in health surveys of low-income, inner-city populations.

An Examination of Internet Users in Iraq

By Kevin Tillmann, John D. Willingham, and Brittany Shanks D3 Systems, Inc. Kevin Tillmann, D3 Systems (kevin.tillmann@d3systems.com)

Internet access is increasing in the developing world, particularly in the Middle East. Past research (Feld and Shanks 2009) indicates that a significant proportion of the young, single, disaffected male Internet news using population in the Middle East is located in Iraq. These Iraqi audiences are often less-educated members from lower socio-economic strata, rather than the young, single males who are well-educated and reside in the upper classes of their countries. They frequently express a negative outlook on their future, and they harbor very strong feelings regarding the withdrawal of US troops from Iraq and the conflict between Israelis and Palestinians.

This paper continues this analysis of the Iraqi Internet news-using audience. It establishes a definitive measurement of Internet news use across Iraq's national population and provides analysis of Internet use by various attitudinal and demographic segments of the Iraqi population. It especially focuses on the disaffected segment, with a sufficient sample size to support analysis of measurements of this population, and makes recommendations for policy-makers to approach communicating with and defusing this threatening population.

The D3 Systems study was conducted face-to-face by native Iraqis interviewers nationwide in Iraq between November 23rd and November 26th, 2008. Total n=10,800 with a sampling margin of error +/- 1%. The survey was a national survey of the general public throughout the country.

The Buffalo Longitudinal Study: Finding Respondents After a 15-year Lapse.

Kelly S. Marczyński, Center for Health and Social Research at Buffalo State College (marczyks@buffalostate.edu); William F. Wieczorek, Center for Health and Social Research at Buffalo State College (wieczowf@buffalostate.edu)

The Center for Health and Social Research (CHSR) at Buffalo State College is currently conducting a follow-up study on a sample of 625 young men aged 16-19 who were originally interviewed from 1991 through 1996 (three times, 18 months apart). Each time a subject was interviewed, he supplied a list of persons who could be contacted to assist in locating him if future funding became available for follow-up. Each subject's social security number was obtained as an additional aid. Since then, with the advent and subsequent popularity of the Internet and cell phone usage, a wide array of potential ways to locate subjects has emerged. The purpose of this study is to examine which of those methods has worked best in this population, and how that translates for tracking in other populations. To date, of the 625 original subjects, 424 are being tracked, 193 (45.5%) of whom have already completed their follow-up interview. A large proportion of subjects (44%) were found using Accurint (a tool available on the Internet to aid in locating people), while about a third were found using the list originally supplied by the respondent. Seventeen percent of the respondents were still living at the same address or phone number as in the initial study. Examination of the breakdown of how the participants who have completed the study thus far were located, along with detailing of the various methods utilized to locate respondents when the initial contact information they had supplied was outdated will lead to a structured outline one can utilize to attain a high follow-up participation rate.

A Survey of Jurors' Perceptions of Internet and Social Media Technology Use in the American Trial Process

David Flores, The Grant Sawyer Center for Justice Studies (dflo30@gmail.com)

This research examines the attitudes and behaviors of a specific group of the American public: citizens who have served as jurors in legal trials.

A pressing issue within contemporary American courts centers on jurors' use of internet and social media technologies. Both jurors' use of the internet to independently research trial-relevant issues, and their use of social media to communicate with others outside the jury, violate central legal standards and pose serious threats to the fairness and integrity of the trial process.

The frequency of incidents has initiated a growing debate about remedial measures, and a wide array of reforms has been proposed. Little is currently known, however, about jurors' perceptions, attitudes, and behaviors in this domain, despite the fact that knowledge on this front may be critical to the successful development of remedial measures.

In order to remedy this deficit, a sample of 156 former jurors participated in a survey designed to examine their perceptions of their experiences during jury service, and their attitudes and behaviors regarding the relationship between internet and social media technology and the trial process.

Results confirmed the gravity of the problem. Nearly two-thirds (63%) of respondents indicated they would be willing to use the internet to research case-relevant issues if they felt it would aid their decisions. A substantial proportion (68%) also indicated perceiving that other jurors may be engaging in such behaviors.

Regression analysis revealed several significant predictors of jurors willingness to utilize the internet to research or communicate about the case. These included: opinions that insufficient information was

provided during the trial, internet and social media use habits, perceptions that others were likely engaged in similar behaviors, likelihood of act being detected, and attitudes that these behaviors were in the interest of justice ($p < .05$).

Implications of the findings for legal policy and procedure are discussed.

Disclosure, Citation and Reuse

Marc Maynard, The Roper Center for Public Opinion Research (marc.maynard@uconn.edu); Lois Timms-Ferrara, The Roper Center for Public Opinion Research (lois@uconn.edu)

The Roper Center has for many years been a major repository for commercial, media and academic survey research data, supporting the documentation, preservation and access to thousands of survey data sets. This poster will: 1) review current disclosure guidelines, including citation information and standards, in the context of materials stored by the Center, including those outlined by AAPOR, NCPP, a National Academy of Sciences Report entitled "Ensuring the Integrity, Accessibility and Stewardship of Research Data in the Digital Age," as well as various publications in fields that frequently analyze opinion data; 2) communicate efforts to expand and modify the Roper Center's own standards for accepting survey data and documentation; 3) define the benefits, challenges and concerns in pursuit of disclosure; and 4) introduce a prototype system to provide open/public access to methodological metadata to support AAPOR's Transparency Initiative.

Is a Border Always Just a Border? U.S. Public Opinion on the Canadian and Mexican Borders **Timothy B. Gravelle, Gallup (tim_gravelle@gallup.com)**

Increased security measures on the United States' borders with Canada and Mexico as part of the Western Hemisphere Travel Initiative, as well as recent statements by American policymakers, have pointed to greater uniformity in U.S. border policy. This marks a shift in U.S. policy and practice, which has historically viewed and managed the Canadian and Mexican borders differently. But what does the American public think – are the Canadian and Mexican borders the same or are they different? This paper analyzes recent data from a Gallup Panel survey on border issues to determine whether the American public perceives differences between the Canadian and Mexican borders. It asks whether Americans' views of Canada and Mexico differ with respect to the need for increased border security, the security threats posed by the Canadian and Mexican borders, and the importance given to economic, security and other ties to Canada and Mexico.

Attitudes of Buddhists, Protestants, and Catholics towards Other Religious Groups in Korea **Jibum Kim, NORC (kim-jibum@norc.org); Jeong-Han Kang, Yonsei University (jkan9130@gmail.com); Yongmo Lee, Konkuk University (ymlee@konkuk.ac.kr)**

In Korea, East and West religion coexist: 23% Buddhists, 18% Protestants, 11% Catholics, and 1% Other religions with 47% non-religious. Recent trends of religious distribution show continuous growth of Catholic population but little change in Buddhists and Protestants populations. Since the attitudes toward religious groups may have an implication for the religious growth of a specific religion and peaceful coexistence of different religious groups, we explore if positive attitudes toward other religions vary by religious group. Using the 2008 Korean General Social Survey (KGSS), we found that 31% of the public has positive attitudes towards Protestants, 52% for Catholics, 53% for Buddhists, 12% for Muslims, 11% for Hindus, 12% for Jews, and 26% for atheists or non-believers. In the multivariate logistic regression, all religious groups have positive attitudes toward their own religion. These positive attitudes for their own

religion are reinforced by stronger religiosity for Buddhists and Catholics, but not for Protestants. An interesting finding is that Buddhists and Protestants have mutual negative attitudes, and Protestants' negative attitudes toward Buddhists are much higher for those who have strong religiosity. While Buddhists do not show significantly positive attitudes toward Catholics, Catholics show positive attitudes toward Buddhists. Also, Catholics do not show significantly positive attitudes toward Protestants, but Protestants show significantly positive attitudes to Catholics. Our findings may suggest the continued growth of Catholic population and the uneasy relationship between Buddhists and Protestants.

You Want To Measure What? Reasons Why Households Refuse Participation In Computer Research

Jeff Scagnelli, The Nielsen Company (jeff.scagnelli@nielsen.com)

The past few years have seen a dramatic migration of media content from traditional modes like TV and newspapers to the Internet. With this movement comes a need to measure and understand the usage of this content. This research was designed to gain insight into the issues and concerns that individuals have with participating in computer research, in order to improve cooperation and has led to improvements that will be implemented in the future expansion of computer research within the Nielsen People Meter sample.

Participation in computer research involved the installation of software onto a cooperators computer. The software collected URL information, software usage on the computer and the presence of any hardware. The research was conducted during the recruitment of 375 households that also participated in the Nielsen TV Ratings People Meter sample. This sample included two subgroups; new households that were recruited for the TV ratings and computer participation at the same time and households already participating in the TV ratings who were then approached for computer participation. Data collection occurred through the use of a refusal report form completed by field staff members when a household refused participation in computer measurement, regardless of their TV participation. The areas that were addressed included; household demographics, presence of different household members, reasons for refusal, amount of information communicated and suggestions for improvement.

Identifying Measurement Error Introduced by Harmonization of Ancestry Data.

Christopher D Ward, University of Michigan (cdward@umich.edu); Jeremy Albright, University of Michigan (jeralbri@umich.edu); Felicia LeClere, University of Michigan (fleclere@umich.edu); Pamela Smock, University of Michigan (pjsmock@umich.edu); Lynette Hoelter, University of Michigan (lhoelter@umich.edu); Peter Granda, University of Michigan (peterg@umich.edu)

A significant body of literature exists on the measurement of race over time, but a much smaller body of literature exists on the measurement of ancestry – the geographical origin of one's lineage – over time. An outgrowth of the Integrated Fertility Survey Series (IFSS), this paper examines problems inherent in and solutions to the measurement of ancestry over time in the Growth of American Families (GAF) studies of 1955 and 1960, the National Fertility Survey (NFS) of 1965, and the National Surveys of Family Growth (NSFG) of 1973, 1976, 1982, and 1988.

Ancestry is important to the study of fertility and family growth. It has been used to predict incidence of breastfeeding, estimate contraceptive failure rates, and examine trends in teenage pregnancy. However, ancestry has been measured inconsistently over time in the GAF, NFS, and NSFG studies, which poses problems for analysts of harmonized data.

Hahn and Stroup (1996) put forth a set of seven primary criteria for measuring ethnicity and ancestry over time, and we demonstrate how these criteria are applied to the harmonization of ancestry variables in the GAF, NFS, and NSFG studies for the IFSS project. We further show that although harmonizing these data renders more-efficient analyses, it introduces measurement and specification error. This paper evaluates the tradeoffs between the improved ease of analysis and the measurement error introduced by harmonization. To examine these tradeoffs, we develop a model to assess the criterion validity of harmonized ancestry variables using multinomial logistic regression.

Ideological Coalitions in the U.S.: The Case Of Liberalitarianism.

Zeljka Buturovic, Zogby International (zeljka@zogby.com); Katherine Schwalbe, Zogby International (katy@zogby.com); Grace Ren, Zogby International (grace@zogby.com)

In recent years several writers have speculated about the possibility of a liberal-libertarian political coalition, dubbed “liberalitarianism”. Here we report results of several surveys that lead us to believe that such coalition is unlikely to materialize in the near future. In all of those surveys we asked our respondents to describe their ideology qualitatively (progressive, liberal, moderate, conservative, very conservative, libertarian) and quantitatively (by choosing a number on 1-9 left-right scale). According to our data, around 2% of likely voters self-identify as “libertarians”. Our surveys have consistently shown that libertarians see themselves on the right-side of the political spectrum. For example, according to a September 2009 Interactive survey of 4279 American adults, an average progressive scores 1.77, an average liberal 2.72, an average moderate 4.78, an average conservative 7.09, an average very conservative 8.54 and an average libertarian 6.57 on the 1-9 left-right scale. In addition, our July 2009 Interactive survey of 41179 American adults showed both conservatives and libertarians considered economics (as opposed to social issues, foreign policy or the environment) to be the central issue determining their ideology. Finally, our December 2008 study of 4835 adults showed a very close alignment between economic beliefs of libertarian, conservative and very conservative adults and a great gap between economic beliefs of libertarians and progressive and liberal adults. We propose that the most likely reason for the long-running ideological coalition between libertarians and conservatives is the shared centrality that their economic views have for them as well as a very strong alignment in their understanding of economics. Taken all together, we propose that the probability of a coalition between liberals and libertarians in the nature future is slim.

The Role of Green Colleges in Society

Rashna Ghadialy, The National Opinion Research Center (NORC) (ghadialy-rashna@norc.org); Jesse Rude, The National Opinion Research Center (NORC) (rude-jesse@norc.org); Lekha Venkataraman, The National Opinion Research Center (NORC) (venkataraman-lekha@norc.org)

Historically, colleges are institutions that not only impart knowledge but also prepare young people to make independent decisions. A prerequisite for independent decision-making is an awareness of the alternatives and a critical attitude toward that which is taken for granted. As “going green” moves from eclectic to chic and more mainstream, “green” higher educational institutions can play an important role in preparing young adults for tomorrow’s ecological challenges by helping them make more environmentally sustainable choices. Sustainability, which can be applied to communities, businesses, and daily practices, provides a solution to further social and environmental degradation.

Sustainability is still a relatively new concept and different organizations develop and use varied definitions and indices of “green.” We examine a popular index of sustainability developed by the Sustainable Endowments Institute (SEI), a nonprofit organization that conducts research on sustainability in college campus operations and endowment practices. In its annual College Sustainability Report Card,

SEI assigns grades to schools based on: administration, energy use and emissions, endowment transparencies, food and recycling programs, green building, investment priorities, shareholder engagement, student involvement, and transportation.

We draw on data from the Integrated Postsecondary Education Data System (IPEDS) to compare the institutional and demographic features of high-scoring “green schools” with those of schools rated poorly on the report card. Using IPEDS data, our research explores what systematic differences may exist between schools with relatively higher and lower sustainability grades. Specifically, we ask: What is the profile of a “green school”? Are schools with fewer financial resources significantly less likely to engage in sustainability practices? And what connections may exist between an institution’s sustainability efforts and its academic outcomes? We further speculate how differences in institutional sustainability may influence student attitudes and future enrollments.

Balancing Data Quality with a More Efficient use of Collection Resources: Experiences from Statistics Canada CATI Surveys

Amélie Lévesque, Statistics Canada (amelie.levesque@statcan.gc.ca); Ann Lim, Statistics Canada (ann.lim@statcan.gc.ca); Christian Nadeau, Statistics Canada (christian.nadeau@statcan.gc.ca); Owen Phillips, Statistics Canada (owen.phillips@statcan.gc.ca)

In an effort to reduce costs and respondent burden, Statistics Canada adopted in 2007 a maximum number of call attempts of 25 for social surveys that collect data through Computer Assisted Telephone Interview (CATI). In the time since its implementation, a number of surveys have successfully petitioned for temporary exemptions from the call limit, often presenting as argument the impact such a restriction would have on response rates. In the summer of 2008, a coordinated effort was made among a number of surveys, including the Survey of Labour and Income Dynamics and the Youth in Transition Survey, to evaluate the impact that a cap of 25 call attempts would have on the quality of the resulting estimates. Non-response was simulated in the existing samples assuming a cap of 25, survey weights were recalculated and a large number of key estimates were compared. The study found that, in general, there was a non-negligible impact to the quality of estimates, and recommended for these two surveys an indefinite limit of 40 call attempts. The presentation will examine in greater detail the results of the study and provide additional evidence in support of the higher cap.

Creatures in the Classroom: Attitudes and Beliefs about Fantastic Beasts, Extraterrestrials, and Creationism Among Education Majors.

Susan Carol Losh, Florida State University (slosh@fsu.edu); Brandon Nzekwe, Dept of Educational Psychology & Learning Systems (nzekweb Brandon@hotmail.com)

Pseudoscience support (cognitions about material phenomena that claim to be ‘science,’ yet use non-scientific evidentiary processes--authoritative assertion, anecdote or unelaborated ‘natural’ causes) is a continuing issue for public understanding of science research. Horoscopes abound; alternative medicine can kill; and battles persist over science curricula. Pseudoscience attitudes ill equip citizens for a society requiring increasingly informed science judgment.

Pseudoscience studies typically address evolution attitudes among liberal arts undergraduates or general public adults. We compare education majors with national adult data on several pseudoscience dimensions. Educators are disproportionately influential and more is expected of teachers than of other adults partly because the public highly values them as societal contributors (Pew, 2009). Further, elementary educators usually teach all fields, providing a foundation for future science literacy.

545 female and 120 male upper-class students in required education courses at a state university completed surveys during 2007. Half were in elementary education; 3.5% were science education majors. We compare them with Associate or baccalaureate adults ($n = 1139$) from the National Science Foundation Surveys of Public Understanding of Science and Technology 2001 and 2006.

Both samples answered science knowledge items from late elementary school. Pseudoscience dimensions for education majors included: “fantastic creatures”, e.g., “Bigfoot”; psychics; astrology; extraterrestrials; evolution; Biblical creation and “Intelligent Design”. NSF Surveys adults answered astrology, psychics, extraterrestrials, and evolution items.

Both groups were similarly knowledgeable—and similarly accepted pseudoscience, e.g., 42% of college-educated adults agreed or were unsure that some UFOs are extraterrestrials, compared with 40% of education majors. Excepting science education students, most rejected evolution, endorsing some form of creationism. Analyses include N-way ANCOVAs and multiple regression using gender, ethnicity, major, science knowledge, and grades as predictors. These predicted evolution support, but not pseudoscience items, e.g., astrology.

Policy implications of these results for education and presentational media are discussed.

The Health Care Confidence Index: Development and 9-Month Findings.

Peter Graven, University of Minnesota (grave165@umn.edu); Michel Boudreaux, University of Minnesota (boudr019@umn.edu)

Public opinion on health reform has a dynamic relationship with attitudes on present and future access to care. Furthermore, changes in policy likely shape opinions on access. To better understand the public's confidence in current and future access to health care, we developed an index that tracks health care confidence. Survey questions on access and attitudes were modified from existing instruments and added to the University of Michigan's Survey of Consumers, a monthly, on-going telephone survey of 500 people in the contiguous U.S. Exploratory factor analysis identified nine items that pointed to a single latent construct and Cronbach's alpha was .86 on baseline data. These nine survey items were aggregated into an index using a method similar to that used by the Consumer Sentiment Index.

In addition to describing the development of the Health Care Confidence Index, we present 9 months of data. We demonstrate the index's association with other survey items, including employment, income, health insurance coverage, general self-rated health status, and opinions on the importance of health reform. We also make use of the survey's rotating panel and present score changes of people who were interviewed on multiple occasions. The index showed a decline in confidence in future access over the summer of 2009, but increasing confidence in the late fall of 2009. We demonstrate that these gains cannot fully be explained by changes in economic well-being and argue that this trend and the index's subsequent behavior are related to health reform developments. We also suggest ways that other researchers could use the index for health policy and public opinion research.

Implementation of Responsive Design for CATI Surveys at Statistics Canada

Milana Karaganis, Statistics Canada (milana.karaganis@statcan.gc.ca); François Laflamme, Statistics Canada (francois.laflamme@statcan.gc.ca)

With the recent emphasis on the analysis of collection process data, paradata research has been focused on a better understanding of data collection processes leading to the identification of strategic improvement opportunities. For Computer-Assisted Telephone Interview (CATI) surveys, research

findings have indicated that the same data collection approach does not work effectively throughout an entire data collection cycle, stressing the need to develop a more flexible and efficient data collection strategy. Over the last two years, the relationship between the quality, cost, productivity and responding potential of outstanding cases over the course of collection has been investigated. Additional tools have been developed to better assess and monitor progress, quality and performance during collection and to allow the development and implementation of Responsive Design (RD) strategy for CATI surveys at Statistics Canada. An RD approach requires ongoing monitoring and analyses of collection progress against a predetermined set of quality, productivity, cost and propensity indicators. The RD objectives are to identify critical data collection milestones and adjust collection strategies correspondingly to allow the most efficient use of available resources remaining. This paper provides an overview of the RD strategy for CATI surveys at Statistics Canada and describes the implementation of the approach for the Households and the Environment Survey (HES), including new tools and indicators that were used to monitor collection and identify data collection milestones.

Country of Origin Representation Among Hispanics Panelists. Arianne Buckley, Arbitron Inc. (arianne.buckley@arbitron.com)

Arbitron has developed an electronic Portable People Meter (PPM) that automatically detects audio exposure to encoded radio signals. We ask panelists to wear this device everyday from when they rise to when they retire to measure their media exposure. Our PPM panels consist of households that vary demographically within over thirty metropolitan areas in the United States. Households are recruited using RDD and address-based frames. Arbitron continually evaluates the composition of the panels to ensure that they closely represent the markets' Census populations.

Recently, there has been a growing interest in how well Hispanic sub-groups are represented within our panels, specifically with respect to panelists' country of origin. To assess the extent to which country of origin distributions among our Hispanic panelists compare to Census distributions, Arbitron tested asking the country of origin question during panelists' regularly scheduled characteristic update in three markets with large and diverse Hispanic populations. The panelist's characteristic update asked a series of demographic question, including the Census country of origin question. Because country of origin can be a sensitive issue, we first wanted to ensure that the addition of this question would not generate any negative responses and impact panel participation. Second, if we were able to capture this information, we wanted to know how well our panels represent the country of origin breakouts within the market populations.

This presentation will discuss the analyses Arbitron has conducted on Hispanic representation. It will provide the details of our country of origin test among Hispanic panelists that resulted in the addition of the country of origin question into our methodology.

Exposing Undergraduate Students to the News: A Comparison of Online and Print Newspaper Reading Assignment in an Undergraduate Course
Leora Lawton, TechSociety Research (lawton@techsociety.com)

An informed citizenry is a core ingredient for a democratic society, but even as issues become more complex and weighty, people are reading newspapers less and less. This phenomenon is especially pronounced for people under 30. While some point out that the Internet has increased access to news coverage, others argue that the quality of news has declined. In this paper, I present results from a college classroom quasi-experiment pilot study (n=52), where students were given a pre-test about news media consumption, then given a 2 week assignment to read a daily newspaper, both online and print,

write about their experience, and then students were given a post-test, using both qualitative and quantitative data. The majority preferred online news because it is free, easy to access, updated frequently and has links to related stories. Those who preferred print enjoyed the experience and ritual of sitting and reading, perhaps in a café, or sitting comfortably at home. Even those who preferred online noted that the depth of news read was greater in print. Students thought that well-defined limited newspaper assignments should be a regular part of a college education.

The Impact of Personal Appeals in Customer Experience Measurement

Joe Cardador, Service Management Group (jcardador@servicemanagement.com); Mary Sheridan, Service Management Group (msheridan@servicemanagement.com)

Organizations interested in capturing customer feedback often print invitations for customers to share their experience on customer receipts. The message directs customers to a web URL or phone-number where they can take a survey about their experience. Because not all customers are aware of these surveys or invitations, survey sponsors are often tempted to use employees to point out survey invitations to customers and request their participation. The logic is that employees have the most engagement with the customer and are therefore in the best position to increase survey participation. However, using employees to highlight the survey invitation creates a problem in that the subject of the survey itself is the performance of and the experiences and services delivered by the employees. As such, employees may be motivated to point out the survey more often to customers who they perceive had a positive experience thus increasing selection bias. In other cases, employees may actually make personal appeals for customers to provide high ratings on the survey, independent of the experience or service they received. This study looks at the impact on survey scores and data quality when employees point out the survey invitation to respondents and when they ask respondents to provide favorable ratings. Recommendations on how to best involve employees in the survey process while minimizing selection and response bias are provided.

Predicting the Index of Consumer Sentiment when it isn't measured

David Fan, University of Minnesota (dfan@umn.edu)

Poll data are typically collected only after real-world developments cause major changes in opinion about a topic. Thus survey values are missing during times of slow opinion change. This irregular timing prevents studies during crucial transitions from relative stability to rapid change.

The missing opinion values can be predicted accurately from persuasive information alone using the ideodynamic model (Fan and Cook, J. Math. Sociol. 27:1-23, 2003). This model has similarities to the linear, autoregressive models used more frequently for time trend analyses.

Both models predict opinion in a series of iterative steps with opinion at a later time computed from prior opinion and persuasive information, both measured with error. In linear models, errors in predicted opinion grow over time if the input is prior, predicted opinion. Consequently, linear models predict future opinion from empirically measured past opinion at every calculation time.

In contrast, the ideodynamic computation can be based on opinion predicted from earlier times because errors from prior predictions decrease with time to give an opinion time trend with errors that stabilize to a constant variance. Therefore, ideodynamics can continue to make accurate predictions indefinitely from persuasive information alone.

This paper uses the methodology of the Fan and Cook paper cited above to demonstrate this ability. The predictors were press coverage scores, and the dependent variable was opinion in the form of the University of Michigan's Index of Consumer Sentiment. The test was to estimate parameters following the removal of monthly Index values from an initial 7 years, a middle 7 years and a final 7 years of a 23-year time trend. The estimated parameters were used to predict time trends in the 7 year gaps, and the predictions were compared to the removed Index values. The predictions in the gaps maintained R-squared values of approximately 0.75.

The Emergence of Nanotechnology Knowledge Gaps: Differences in Knowledge Across Education Levels and Media Exposure.

Michael A. Cacciatore, UW-Madison (mcacciatore@wisc.edu); Dietram A. Scheufele, UW-Madison (Scheufele@wisc.edu); Elizabeth A. Corley, Arizona St. University (corley.elizabeth@gmail.com)

Nanotechnology is one of the most groundbreaking technological breakthroughs in recent memory. Nonetheless, research on the opinion dynamics surrounding this emerging technology and its potential risks and benefits is still in its infancy. Despite increased spending efforts aimed at communicating with the U.S. public about nanotechnology, initial research seems to suggest that nanotechnology knowledge levels among the public have remained relatively static over the past few years. However, most of this research has examined changes in nanotechnology knowledge for the public as a whole or in simple cross-sectional studies as opposed to examining the differences across diverse sets of publics and across multiple data collections.

In this study we take a more detailed approach by examining U.S. public knowledge levels across different levels of education and media use. In particular, we explore changes in knowledge levels among nationally-representative samples in 2004 and 2007 for different groups based on education levels and media exposure using data from two nationally representative telephone surveys.

First, our analysis finds widening nanotechnology knowledge gaps between 2004 and 2007 for those U.S. respondents with low levels of formal education and those respondents with high levels of formal education. Our results show that while respondents with at least a college degree have experienced an overall increase in nanotechnology knowledge between 2004 and 2007, respondents with less than a high school degree have actually experienced a decrease in nanotechnology knowledge levels over the same time period. Second, we find that exposure to the internet significantly improves knowledge levels of those with the lowest levels of education. At the same time, there is a less positive impact of internet exposure on those with higher levels of formal education. We conclude by discussing the implications of our findings.

Effects of Changing Citizenship Status on Non-response

Brianna Groenhout, NORC (groenhout-brianna@norc.org); Deborah Harper, NORC (harper-deborah@norc.org); Lance Selfa, NORC (selfa-lance@norc.org)

Previous research has shown that citizenship status has a significant relationship to survey non-response and item non-response (e.g. Hoffer, Hess, and Lisek, 2007, Paper presented at the 62nd AAPOR conference). Drawing on data from the Survey of Doctorate Recipients (SDR) and the Survey of Earned Doctorates (SED), this paper adds a new dimension to the literature, exploring whether change in citizenship status over time interacts with propensity to respond at both the unit and item level. Citizenship status is broken into four categories: native born U.S. citizen, naturalized U.S. citizen, non-U.S. citizen with a permanent resident visa, and non-U.S. citizen with a temporary visa. The SDR is a longitudinal survey conducted biannually that follows the educational attainment and work patterns of

individuals who have received doctorates from U.S. institutions in the fields of science, engineering, and health, and are living in the U.S. at the time of the survey. The 2003 SDR collected detailed information on citizenship, such as reasons for coming to the U.S., type of visa held, and when green card status was obtained. The SED is an annual census and survey of research doctorate recipients from U.S. institutions. The Doctorate Records File (DRF) is the cumulative file of all completed years of SED data, and serves as the sampling frame the SDR. The SED collects citizenship status and post graduation plans, including type of plan and location. Using combined SED and SDR data, we will model the relationship between unit and item non-response, and citizenship status. We will control for key demographics, reported post graduation plans at the time of doctorate receipt, length of stay in the U.S. for non-U.S. citizens, educational history, and employment history. We will explore whether and how a change in citizenship status affects unit and item level non-response.

Do Deliberative Polls Educate Better Citizens? - Comparing Two DPs in a Local Township in China.

Rui Wang, Stanford University (wrui@stanford.edu); Alice Siu, Stanford University (asiu@stanford.edu)

Critics of deliberative democracy say that the public is too confused, inconsistent, and ignorant to be worth consulting about policy issues. Advocates of the Deliberative Polling believe that the public is capable of making informed decisions driven by information and by the consideration of competing arguments. Moreover, the process of deliberation can even serve as an educational program to produce better citizens who are more active and competent in political participation and policy making. In February 2009, the local government of Zeguo Township in Wenling City, China conducted its fourth Deliberative Poll (DP). Participants of the 2008 DP were invited again to deliberate on the Township's 2009 financial budget allocation. The two DPs, targeting the same citizens, provide a good data source to track the opinion changes and test the educational effect of the DP. This paper analyzes and compares their opinions and knowledge in the two DPs, as well as their evaluations of the two DP events. Since the local government reported the incomplete and complete projects in 2008 to the public before the 2009 DP event, this paper also explores whether people's opinions about government's implementation based on the 2008 DP results would have an impact on the effectiveness of the educational function of the DP.

Lessons Learned from Years of Fielding Employer Surveys.

Heidi Whitmore, NORC (whitmore-heidi@norc.org); Jeremy Pickreign, NORC (pickreign-jeremy@norc.org)

Sampling

We have found that completion rate varies across multiple factors, with previous survey participation and firm size being most important. Firms with previous survey participation are three times as likely as new firms to respond, while the very largest and smallest are least likely.

Questionnaire

We have found that health insurance is a complex topic even for benefits managers, and that the instrument must have great detail and validity checks. Many respondents are uncertain as to which types of health plans they offer, and we have created a series of questions to accurately determine this.

Fielding

We do extensive "pre-calling" to identify the name and contact information of the respondent. This also increases the odds of their receiving the advance letter (outlining the content), which is very helpful in increasing participation. If the respondent did not receive it, we offer to re-send. Among firms designated

for pre-calling, contacted firms may have a response rate 50% higher than non-contacted firms. We estimate from one of our more recent surveys an additional 7% response rate.

We use telephone as the mode. Experience has shown that extensive in-person follow up is needed to achieve our target response rate. We make an average of nearly 40 calls per firm in an effort to complete. We also spend considerable effort tracking down the right person in the firm to speak with.

As part of the fielding, we conduct refusal conversion on an eligible portion of the sample. Among firms eligible for refusal conversion in one of our more recent surveys, 9% were converted to a complete (7% of the completed sample).

Incentives

We have found that respondents are very interested in obtaining a copy of the results, and that offering this is very helpful in obtaining cooperation. We also have found the opportunity to benchmark is important.

Better Late Than Never: The Impact of Callbacks on Response Rate.

Larry Luskin, ICF Macro (lawrence.a.luskin@macrointernational.com); Todd Hamlin, ICF Macro (todd.hamlin@macrointernational.com); Michael Ross, ICF Macro (michael.s.ross@macrointernational.com)

Many of us are continuing to use phone-based surveys as a means to gather information from our sample populations. Since 2005, ICF Macro and the Internal Revenue Service have been using a phone-based survey to conduct the Taxpayer Advocate Service (TAS) Customer Satisfaction Survey. The use of callbacks has been an integral part of the data collection process to maximize response rates. How effective are these callbacks, not only to increase overall response rates but to minimize the potential for non-response bias? How much non-response bias is actually present, what is the nature of this bias, and how might it affect our results? To answer these questions, we conducted a non-response bias study that integrated three different techniques: (1) frame-to-sample variable comparisons, (2) respondent subgroup response rate comparisons, and (3) an examination of respondent differences as a function of the follow-up effort (e.g., callbacks). We will also address such questions as the incremental impact of callbacks, and the impact of late responders and how their satisfaction scores compare to early responders. These findings will serve as the basis for creating a predictive model of how non-responders may impact survey results.

The Effects of Respondent Fatigue

Tricia McCarthy, National Opinion Research Center (mccarthy-tricia@norc.org); Lauren Seward, National Opinion Research Center (seward-lauren@norc.org); Kymn Kochanek, National Opinion Research Center (kochanek-kymn@norc.org)

The number of contacts required for interview completion with respondents in a longitudinal survey can be extremely varied. Some respondents may need a minimal number of contacts throughout all their rounds of participation, while others require many contacts and great effort put forth by the field interviewer in order to complete an interview. The purpose of this paper is to look at the average number of contacts perfect responders, or respondents who have completed each round of interviewing, and converted non-interview respondents (NIR) may require, as well as variables that might affect this number, such as education, income, and gender. The sample will come from the National Longitudinal Survey of Youth 1997. The NLSY is conducted annually and is currently in its 13th round of interviews with the same sample of almost 9,000 individuals, gathering data on factors regarding economics,

education, and health. The sample looked at in this analysis will consist of all respondents who completed an interview in Round 9 of the survey. The data that will be analyzed for these respondents will come from Rounds 10-12. This will ensure that each respondent starts with the same returning status in Round 10 data. From this sample, we can then look at the average number of contacts for perfect responders, for respondents who have been an NIR at one point in the past and for respondents who are converted NIRs. This sample will exclude any respondents who have been in the military or imprisoned during the course of the study thus far, since these cases can require special efforts to clear security measures or gain contact information. Further insight into the variables that can affect respondent fatigue and interview completion can help future surveys better understand the realities of contacting respondents and the amount of work cases might need.

AAPOR Concurrent Session D

Friday, May 14, 2010

4:15 p.m. - 5:45 p.m.

Database Linking & Other Techniques for Improving Data Quality

Demonstrating the Multi-level Integrated Database Approach for Improving Survey Research

Tom W. Smith, NORC (smitht@norc.uchicago.edu)

MIDA has the potential to advance social-science research in general by notably improving survey-research methodology. Moreover, it does so by drawing on one of the major societal changes in recent decades, the development of large-scale, computerized databases that hold extensive information about individuals, households, neighborhoods, and other societal units.

Methodologically, it should help to increase response rates, allow for a much more comprehensive assessment of non-response bias, and facilitate the calculation of weights and imputations to adjust for the detected non-response bias. Besides providing for a general approach to deal with non-response, it will in particular permit the testing of several prominent theories and hypotheses explaining non-response: social disorganization theory, social isolation theory, overextension theory, structural impediments, etc. The auxiliary data from the databases will permit an examination of general, non-response models. Substantively, MIDA will improve analysis by easily and automatically making multi-level, contextual variables as ready for analysis as data directly collected in surveys. As the list of examples cited above attest, geographic context has notable impacts on many aspects of people's lives. The contextual data from sampling frames and augmented from multiple databases will provide a rich, contextual array of data for analysis across scores of central substantive topics.

Evaluation of Potential Benefits, Costs and Risks in the Integration of Survey and Administrative Data on Consumer Expenditures

John L. Eltinge, Bureau of Labor Statistics (Eltinge.John@bls.gov)

The U.S. Consumer Expenditure Surveys currently collect information through two household-survey instruments involving, respectively, a quarterly recall interview and a diary. For some subpopulations and some types of expenditure, one could consider supplementation of current household survey information with data from administrative records. This paper provides a general framework for evaluation of the potential benefits, costs and risks associated with this prospective use of administrative data. Four points are of primary interest. First, extension of standard estimation methods for multiple-frame, multiple-mode surveys can provide a basis for integration of the survey and administrative data. Second, evaluation of the quality of the resulting estimators generally will involve the degree of population coverage provided by the administrative records; the quality of the administrative and survey variables; the strength of association between the administrative variables and related survey items; and the unit-level propensities for matching survey respondents with administrative records. Third, the relevant cost structures involve a wide range of fixed and variable cost components related to capture of both survey and administrative data. Finally, the evaluation of risks arising from the use of administrative-record sources should include systemic risks as well as more commonly studied aggregate risks.

Key Words: Administrative-record system; Data quality; Dual-frame survey; Fixed and variable cost components; Multiple-frame, multiple-mode survey; Operational risk; Sensitivity analysis; Systemic risk; Total survey error.

Methods of Linking Different Surveys Together When Respondent Samples Differ

Scott Hershberger, Harris Interactive (shershberger@harrisinteractive.com); John Bremer, Harris Interactive (jbremner@harrisinteractive.com)

Frequently, organizations construct and administer surveys on a number of different subjects, each survey being administered to a partly or completely different sample of respondents. The goal of this presentation is to describe methods for linking these various surveys together, in order to determine how respondents would have responded to surveys that were never administered to them. Due to recent advances in statistical methodology, it is now possible to link different surveys together even when the samples differ among the surveys. The result is that for each respondent, there is complete information about the respondent on each of the surveys. These advances in statistical methodology encompass a number of techniques including a respondent matching algorithm based on Bayesian statistics, methods of linear and equipercenile test equating, and methods of imputing missing data. This presentation uses as an example two annually administered surveys: the first involves television viewing preferences, and the second involves consumer brand perceptions. For both surveys, a proportion of the sample is unique, while another portion partially overlaps both surveys. The goal of the survey linkage is to impute for each respondent answers for all of the survey questions, even when the respondent had not been administered the survey. Survey linkage is an invaluable tool that can be used by researchers, pollsters, and commercial marketing firms to leverage information contained in disparate surveys.

Using Statistical Process Control Charting in the Improvement of Survey Data Quality

David Peng, D3 Systems (david.peng@d3systems.com); Karl Feld, D3 Systems (karl.feld@d3systems.com)

Much of Quality Control within survey research is conducted in similar fashion to acceptance sampling, which is one of the earliest forms of Quality Control within the industrial realm. It consists of taking a sample from a lot, with some quality characteristic inspected, and on the basis of the investigation accepting or rejecting the lot.

Acceptance sampling has become de-emphasized within industry because it is “too late, costly and ineffective.” Instead, Statistical Process Control, using information from the Statistics and Engineering disciplines,” has seen significant research done specifically on using control charting in “real-time” to improve quality. With the advancement of CATI center technologies and CAPI surveying devices, there is also an opportunity to take advantage of real time data to deliver control charting for quality in survey research.

The application of control charting to live interviewer performance data allows management to statistically monitor the performance of their interviewers and projects in a proactive fashion. Instead of performing Quality Control periodically during a survey or after the fieldwork is completed, live data analyzed consistently through these charts allows one to take corrective actions and investigation while the process is ongoing, ultimately resulting in a consistently higher quality end product.

The specific focus of this paper reports research on the application of control charts to various industry standard survey measurement metrics based on sample disposition reporting. Examples of applications

in real, international production environments are provided. Standard AAPOR measurement rates derived from disposition data will be used for demonstration purposes in this presentation. Other quality control measures used in the industry will also be discussed.

Identifying Techniques for Improving Questionnaire Design

Seeing Through the Eyes of the Respondent: An Eye-Tracking Study on Survey Question Comprehension

**Timo Lenzner, GESIS - Leibniz Institute for the Social Sciences (timo.lenzner@gesis.org);
Kaczmirek Lars, GESIS - Leibniz Institute for the Social Sciences (lars.kaczmirek@gesis.org);
Galesic Mirta, Max Planck Institute for Human Development (galesic@mpib-berlin.mpg.de)**

Comprehending the question is the first cognitive process respondents have to carry out when answering a survey question (e.g. Tourangeau 1984). If respondents have difficulties performing this task, they are likely to provide inaccurate answers (Schober and Conrad 1997) and/or apply response strategies that reduce data quality and induce measurement error (e.g. satisficing, Krosnick 1991; breakoff, Galesic 2006). Consequently, it is important to design comprehensible questions that respondents find easy to understand and to answer accurately. From a psycholinguistic perspective, survey question comprehension is undermined by several problematic text features, such as low-frequency words, vague or imprecise relative terms, vague or ambiguous noun-phrases, complex syntax, complex logical structures, low syntactic redundancy, and bridging inferences. Earlier research has shown that respondents require more time to answer questions that include these text features, presumably because the text features induce comprehension difficulties (Lenzner, Kaczmirek, and Lenzner 2009). In this study, we used more direct measures of comprehension difficulty, namely eye-tracking parameters. We recorded respondents' eye movements in the laboratory while they were answering two versions of similar questions in a Web survey. One group (n = 22) received questions which contained one of the problematic text features mentioned above and the other group (n = 22) received control questions which did not contain such a text feature. Our results reveal that questions including problematic text features receive longer and more fixations than control questions without these text features. These effects exist for various types of questions: attitudinal, factual, and behavioral. Hence, we advise survey designers to avoid these text features in order to enhance the comprehensibility of their questions.

Measuring Race and Hispanic Origin: Cognitive Test Findings Searching for "Truth".

Jennifer Hunter Childs, U.S. Census Bureau (jennifer.hunter.childs@census.gov); Rodney Terry, U.S. Census Bureau (rodney.terry@census.gov)

There is no true gold standard for measuring race and Hispanic origin when they are, by definition, self-identification concepts. U.S. Census Bureau staff developed and cognitively tested a telephone interview that will serve as a gold standard to evaluate experimental versions of Hispanic origin and race questions during the 2010 Census. Three methods of measuring race and Hispanic origin were developed and tested during cognitive testing that allowed us to generate what we believe is a robust, though never fully comprehensive, portrait of self-identified race and Hispanic origin. The first method is an open-ended question that allows the respondent to self-identify with any races or Hispanic origins. The second method is a series of yes/no questions aimed at measuring identification with the OMB race and Hispanic origin categories. The final method is a summary measure which attempts to gather the respondent's usual or typical report to race and Hispanic origin questions.

This paper provides the results of three rounds of iterative cognitive testing of this telephone interview conducted with 37 respondents, including members of all of the Office of Management and Budget (OMB) race and Hispanic origin groups (including 10 multiracial respondents and 12 Hispanic respondents). Importantly, we find that there is no single true gold standard for measuring race and Hispanic origin. However, this paper highlights the findings from each round, as well as how and why the questionnaire was modified between rounds to achieve an interview that we believe can reasonably be used as a gold standard.

How Long Did You Spend Completing This Questionnaire? – Methodological Challenges on Measuring the Response Burden in Mail-Collected Surveys.

Serge Godbout, Statistics Canada (serge.godbout@statcan.gc.ca)

Response burden in surveys is a major concern. In order to measure the businesses' compliance costs in economic surveys from Statistics Canada, all Unified Enterprise Survey questionnaires contain a question asking "How long did you spend collecting the data and completing the questionnaire". Even though this "time to complete" variable provides key information to assess the response burden, care must be taken in order to use it properly.

In this talk, we present some conceptual issues related to the "time to complete" question and how the respondents might misunderstand it. We describe the methodology developed to enhance its quality, which includes edit rules, outlier detection and imputation for missing values. Some strategies are proposed to calculate the total time spent completing the questionnaires and to predict it on further cycles, prior to the collection stage. Finally, results coming from survey data are presented.

Happy! Or Can't Get No Satisfaction?: Concept Equivalence in The Measurement of Subjective Well-Being.

Patrick Sturgis, University of Southampton (p.sturgis@soton.ac.uk); Patten Smith, IPSOS-Mori (patten.smith@ipsos.com); Bobby Duffy, IPSOS-Mori (bobby.duffy@ipsos.com); Caroline Roberts, FORS (Swiss Foundation for Research in the Social Sciences) (caroline.roberts@gmail.com)

As the limitations of macro-economic indicators, such as inflation and Gross Domestic Product, as barometers of the health of nations have become increasingly apparent, attention has shifted in recent years to the more utilitarian notion that the ultimate goal of government policy should be to maximize the level of positive affect among the citizenry. This new research paradigm, primarily situated within economics and psychology, seeks to address fundamental questions about the drivers of individual contentment: what are the effects of money, status, relationships, and health on the extent to which an individual is satisfied with his lot? Crucial to this growing research agenda are survey measures of what is generically referred to as 'subjective well-being'. And, while analysts tend to treat these items, if only implicitly, as equivalent measures of the same underlying construct, differences in question wording and response format abound. Particularly notable in this regard is the difference between questions which ask how 'happy' the respondent is and those which require an estimate of the level of 'satisfaction' the respondent feels with her life. Our aim in this paper is to assess the extent to which happiness and satisfaction questions elicit equivalent assessments of subjective well-being and to explore the frames of reference that people employ in formulating a response to them. We do this by randomly assigning respondents to a general population omnibus survey to receive either a happiness or a satisfaction question, both of which employ equivalent response formats. Respondents are subsequently asked to state, in their own terms, what came to mind when answering the question. The verbatim responses are coded to a frame and distributions compared across question type. Conclusions are drawn about the equivalence of happiness and satisfaction questions as indicators of subjective well-being.

Improving Web Survey Quality

Study of Non-Probability Sample Internet Surveys' Estimates of Consumer Product Usage and Demographic Characteristics of Consumer Product Users

David Scott Yeager, Stanford University (dyeager@stanford.edu); Autumn Carter, Stanford University (dyeager@stanford.edu); Hiyabel Tewoldemedhin, Stanford University (dyeager@stanford.edu); Jon Krosnick, Stanford University (krosnick@stanford.edu)

Non-probability sample ("opt-in") Internet surveys are often used to measure how many people use various consumer products and to identify the characteristics of those product users. In this study, we evaluate the degree to which different opt-in panels produce the same results when used for such analyses. Seven opt-in Internet survey companies administered the same set of measures of product usage (each N~1000). We found that (1) the seven surveys varied widely in their estimates of how many people used various consumer products; (2) the market share of some brands, such as Coca-Cola, relative to other brands, such as Pepsi, were notably different depending on which survey vendor one used, and the ranking between these and other brands flipped depending on the vendor one used; and (3) the demographic characteristics of people using various products were sometimes the same across surveys and sometimes quite different. Weighting the data to yield optimal distributions of various demographic variables did not change the variability of survey results on consumer product usage across opt-in surveys. Overall, these results suggest caution before generalizing from a product usage survey conducted with one opt-in Internet panel to the population of all such surveys.

Time Related Inconsistencies in Global Online Panels

Steven Gittelman, Mktg, Inc. (steve@mktginc.com); Elaine Trimarchi, Mktg, Inc. (Elaine@MktgInc.com)

Respondents who are active members in online panels and enjoy the interviewing process, do surveys more often, and stay in the system for years. They have become the core of the online respondent community. In the United States, those who complete over thirty interviews a month are the core of the interviewing community and represent over half of those constituting the panel system.

The average respondent serves on 4.1 panels and completes the lion share of interviews being conducted in the commercial sphere.

Respondent tenure has been recognized as a factor influencing purchasing intent. In 29 studies, representing some 40,000 interviews conducted by Ron Gailey of Washington Mutual during 2006-7, a 30% drop in purchasing intent for WAMU's financial products was noted. As reported by Ron Gailey in 2008, these trends did not coincide with sales report data from the field. Washington Mutual is gone and we will never know if incorrect business decisions were made from this research data.

Many variables logically tie into respondent tenure. Panelists who enjoy the interviewing process can become truly hyperactive. We hypothesize that such respondents will behave differently from those who are new to the system. Indeed if panels aggregate too many such respondents they will no longer represent the population that they were drawn from, no matter how rigorous the sampling effort might originally have been.

During 2009, we surveyed 180 panels in 35 countries focusing on buyer, media and socio-graphic segmentations. In this paper, we will show that respondents, who are in multiple panels and complete surveys frequently, present extremely different buying behaviors. The numbers of these respondents and the impact it has on buying behavior are substantial and outweigh numerous online research quality efforts being considered.

Does Making The Survey Topic More Salient Lead To An Expert Bias? – The Influence of Announcing Survey Topics in E-Mail Invitations for Web Surveys in Online Access Panels on Response Rate and Data Quality.

Florian Keusch, Vienna University of Economics and Business, Austria (florian.keusch@wu.ac.at); Wolfgang Mayerhofer, Vienna University of Economics and Business, Austria (wolfgang.mayerhofer@wu.ac.at); Nicole Weilbuchner, Vienna University of Economics and Business, Austria (nicole.weilbuchner@wu.ac.at); Silke Jungreithmaier, Vienna University of Economics and Business, Austria (silke.jungreithmaier@wu.ac.at)

Personal interest for a survey topic has proved to substantially influence the decision to participate in mail (Dommeyer 1985, Heberlein/Baumgartner 1978, Martin 1994) as well as web surveys (Galesic 2006, Marcus et al. 2007, Porter/Whitcomb 2005). According to the 'Leverage Saliency Theory' of Groves, Singer and Corning (2000) the decision to take part is not only based on the different attributes of the given survey (e.g. topic, sponsor, length etc.) but will also be influenced by how salient those attributes are made to the invitee. Invitations for web surveys in online access panels therefore usually leave this information out to reduce the risk of an expert bias (Bethlehem/Stoop 2007). This study aims to bring forward whether online access panel members use the information about the survey topic in e-mail invitations in their decision on participating in web surveys.

In a preliminary study data about personal interest, attitudes and other psychographic variables of 1.670 panel members were collected. Panelists were then randomly assigned to participate in one of two surveys about different topics (insurances and holidays). The e-mail invitations to these surveys were manipulated. Panelists either received an e-mail with the topic stated in the subject line as well as the invitation text, an e-mail message where only the text referred to the topic or an e-mail without any information about the survey topic. The influence of announcing the topic in the e-mail was measured on response rate and different indicators of data quality (response latency, item omission, answer length to open-ended questions, non-differentiation in grids).

Professional Web Respondents and Data Quality

Frederick G. Conrad, University of Michigan (fconrad@isr.umich.edu); Roger Tourangeau, University of Michigan (rtourangeau@survey.umd.edu); Mick P. Couper, University of Michigan (MCouper@umich.edu); Chan Zhang, University of Michigan (chanzh@isr.umich.edu)

There is concern that professional respondents, motivated more by monetary incentives than the opportunity to provide thoughtful answers, may compromise data quality in on-line surveys. The worry is that such respondents primarily take actions on the "critical path" toward their monetary goal even though valid answers may require additional, "off-path" actions. Our main questions are whether professional respondents are quicker to find the critical path and adhere to it more closely than "non-professionals." We explored this in a web survey in which 2404 volunteer respondents answered two constant-sum questions, i.e., multi-part items whose answers must add to 100% (fourteen internet use categories) or 24 hours (eleven daily activities). Respondents could not advance unless their answers were well-formed (i.e., equal to the target sum). About one third of the respondents could click for a running tally to facilitate well-formedness, placing this action on the critical path. We considered respondents who belonged to

more than 3 volunteer panels to be professional (58% had completed more than 30 web surveys). This distinction mattered. Professionals were more likely than non-professionals to request a tally for the first item but the quality of both groups' responses benefited from the tally: more non-zero, non-missing answers and fewer rounded responses with than without the tally. However, by the second item, professional respondents seemed to learn that providing valid answers was not on the critical path. While professionals again requested the tally more often, the tally only improved the quality of answers for non-professionals: more non-zero, non-missing answers and less time spent on "other" activities with than without the tally; professionals' answers showed no such improvement from the tally. The results suggest that professional respondents are more likely to provide well-formed than valid answers because validity is not on the critical path.

Using Overt and Covert Survey Traps to Maximize Data Quality

Joe Cardador, Service Management Group (jcardador@servicemanagement.com); Melissa Wayman, Service Management Group (mwayman@servicemanagement.com); Mary Sheridan, Service Management Group (msheridan@servicemanagement.com)

Survey researchers using the web for data collection often must contend with respondents who provide poor quality data. Poor data can come from careless responders or fraudulent responders. Careless responders are those who may simply be distracted or fatigued when completing the survey or may respond carelessly and quickly in an effort to expend as little energy as possible in exchange for the survey incentive. Fraudulent responders are those who are interested in intentionally manipulating the survey results or collecting a valued incentive multiple times. This study compares a number of overt and covert methods to uncover fraudulent and careless responding. Overt traps are part of the survey content and are visible to the respondent on the survey, but covert traps are not. Covert traps analyze response characteristics that suggest poor data quality. Our research indicates that overt survey traps were effective in capturing careless responders, but covert traps were best for capturing fraudulent responders. Recommendations on the use of overt and covert survey traps based on our findings are provided.

Interviewer & Measurement Error

Measurement Error in Objective and Subjective Interviewer Observations

Jennifer Sinibaldi, National Centre for Social Research (jennifer.sinibaldi@natcen.ac.uk)

As the collection of paradata gains prominence in survey research, the quality of the paradata has come into question. Paradata routinely collected in face-to-face surveys include interviewer observations of the area and household. These data are often used in propensity models for nonresponse adjustment and/or responsive design decisions. Therefore, important survey outcomes hinge on the quality of these data.

Using the second wave of data collection of a face-to-face panel survey of the Great Britain general population, the measurement error of four interviewer observations routinely collected by NatCen (UK) was estimated. Two of the observations are objective (observing barriers to entry and housing type) while two of the observations are subjective (with the interviewer evaluating the condition of the area and dwelling unit). The design of the data collection also involved an experimental condition, assigning some cases to the same interviewer as the previous year and all others to different interviewers of various experience levels. Therefore, this analysis allows for the assessment of differences in the observations among cases assigned to the same and different interviewers.

Measurement error is first evaluated at the bi-variate level, looking at differences between observations between the two years of data collection. The second stage of analysis uses random effects logistic regression models to predict disagreement in the recorded observations for the same property. Finally, the observations are used to develop nonresponse weights and the impact on the estimates evaluated. Conclusions are made regarding the severity of the differences for each type of observation, factors that may lead to higher incidences of disagreement, and whether the differences have an impact on the final estimates.

Interviewer Observed vs. Reported Respondent Gender: Implications on Measurement Error.
Susan Kenney McCulloch, Joint Program in Survey Methodology (skenney@survey.umd.edu);
Frauke Kreuter, Joint Program in Survey Methodology (fkreuter@survey.umd.edu); **Stephanie Calvano**, Marist Institute for Public Opinion (stephanie.calvano@marist.edu)

Many telephone surveys require, at some point, interviewers to observe and record respondents' gender. Basing these observations solely on a respondent's voice, researchers rely on these "guesses" to inform perhaps: (1) screening to determine eligibility for the study; (2) filtering to assign survey logic and follow-up questions; (3) non-response adjustments; and (4) post-stratification weighting. However, while some although limited, research has been done assessing interviewers' assumptions of respondents' demographic characteristics (e.g., race), no known research has specifically investigated the relationship between an interviewer's guess of respondent gender, the true value (respondent reported) of this important characteristic, and the implications on measurement errors when deviations between observed and reported occur.

This paper presents results of a meta-analysis of 13 independent telephone surveys of United States adult residents. Data collection was conducted by trained interviewers at the Marist Institute for Public Opinion. Across all datasets, we analyze differences between interviewer observed and respondent reported gender. In addition, we examine the relevance of interviewer characteristics (such as gender and race) in understanding discrepancies. A preliminary analysis of these data suggests that approximately 10% of interviewers' gender observations contradict those reported by respondents and that females, in particular, are incorrectly categorized. The paper discusses the implications of such discrepancies on measurement error and suggestions for improved collection of such data.

An Examination of the Quality and Utility of Interviewer Estimates of Household Characteristics in the National Survey of Family Growth
Brady T West, Institute for Social Research (bwest@umich.edu)

Effective methods for repairing nonresponse error are of primary interest to the field of survey methodology, given declining response rates in household surveys of nearly all formats. Post-survey methods for repairing nonresponse error rely on the presence of auxiliary variables on a sampling frame for both respondents and non-respondents, and much methodological work has shown that the best auxiliary variables for repairing nonresponse errors are related to both the survey variables of interest and response propensity. Unfortunately, auxiliary variables having these optimal properties are very rare in survey research practice. In Cycle 7 of the National Survey of Family Growth (NSFG), interviewers performing household screening operations were asked to record their best guesses as to whether there were children under the age of 15 in the household (35,258 guesses by 96 interviewers, prior to the screening questions), and whether the selected respondent was in a sexually active relationship with a member of the opposite sex (13,495 guesses by 94 interviewers, after the completed screening questions). Given that "correct" values on these two indicators can be derived from completed household listings and responses to the main NSFG interview, this study sought to examine the amount of error in

the two interviewer estimates, and the associations of the interviewer estimates with both key NSFG variables and the propensity to respond to the main NSFG interview, given a completed screening interview. Several significant associations were found, suggesting that these interviewer estimates may be useful for repairing nonresponse errors. However, a simulation study shows that the level of estimation error in the NSFG may have a negative impact on potential reductions in nonresponse error. The study concludes with a discussion of estimation techniques used by highly accurate interviewers and future research in this area aimed at improving the quality of the interviewer estimates.

Testing Competing Neighborhood Mechanisms Influencing Participation in Household Surveys
Carolina Casas-Cordero, Joint Program in Survey Methodology (ccasas@survey.umd.edu)

Survey researchers have long recognized that the “social environment” has an influence in the survey participation decision (Groves and Couper, 1998). Influences are identified at the contact and recruitment stage, and they are hypothesized to influence the respondents as well as the interviewers. However, the mechanisms linking neighborhood social processes and individual/household nonresponse have not been adequately tested in the survey literature. Studies to date have only used indirect measures of the ecological constructs of interest (e.g., social disorganization, social cohesion, or collective efficacy), and only a few have adequately modeled the multilevel structure of the data. In addition, the survey literature has not formally elaborated hypothesis about how characteristics of the “physical environment” could also influence the survey participation decision (e.g., physical decay, physical disorder, or mixed land uses). In this paper, I use data from the Los Angeles Family and Neighborhood Study (L.A. FANS) to develop multilevel response propensity models that can overcome these shortcomings. I develop theoretically driven measures of neighborhood social processes and the physical environment using data from the L.A. FANS interview and from interviewer observations of the sampled neighborhoods. I test the influence of the different measures of neighborhood characteristics on contactability and cooperation, controlling for household-level demographic characteristics (available from the screener interview) and for neighborhood-level demographic characteristics (available from census records). This research contributes to our understanding of nonresponse in household surveys in two ways: (1) it is the first test using direct measures of the neighborhood social mechanisms hypothesized to influence nonresponse; and (2) it provides theoretical background - and empirical tests - of how neighborhood physical characteristics could also influence the survey participation decision.

Mechanisms of Error in Housing Unit Listing
Stephanie Eckman, Joint Program in Survey Methodology (steph.eckman@gmail.com)

Our largest and most important national face-to-face surveys use listing to create a frame of housing units for selection, yet we know very little about the quality of these frames. Using a unique dataset with three listings of a nationally representative sample of segments as well as results of debriefings with experienced listers, I report on the housing unit, area and lister characteristics that correlate with overcoverage and undercoverage. I test hypotheses about the mechanisms of lister error—the behaviors that listers engage in that lead to frame error. Preliminary results suggest that listers using the dependent method are likely to carry errors of undercoverage and overcoverage in the initial frame through to the final frame, a phenomenon I call confirmation bias. This finding of confirmation bias calls into question the results of earlier studies of lister error as well as research into the coverage of address databases, which often use dependent listing to create a gold standard frame. My results will help surveys which use listing to understand the quality of their frames and to evaluate the risk of coverage error in their data.

Issues in Survey Non-response

Do Literacy Levels Affect Item Nonresponse In An Interviewer-Administered Survey? An Examination of the National Assessment of Adult Literacy Survey

Ying Wang, University of Nebraska - Lincoln (yingwang@huskers.unl.edu); Kristen Olson, University of Nebraska (kolson5@unl.edu); Jolene Smyth, University of Nebraska (jsmyth2@unlnotes.unl.edu)

With the availability of an address-based frame with a high coverage rate, self-administered mail surveys are becoming an increasingly attractive option for survey designers. A problem arises without the presence of an interviewer in that people with low literacy levels might be unable to answer complex or difficult questions, or may not be able to participate in the survey at all. Yet the relationship between literacy and data quality is surprisingly understudied, as direct measures of literacy require a literacy test. This paper investigates the relationship between respondents' literacy measured through two different methods of literacy assessment and data quality as measured through item-nonresponse. The data are from the 2003 National Assessment of Adult Literacy (NAAL), an in-person area probability survey of 18,186 individuals age 16 and over in the United States (weighted response rate = 60.1 percent). The NAAL measures U.S. adults' ability to read and interpret written materials using three interviewer-administered literacy assessments, providing a strong test for the relationship between literacy and item nonresponse. In this study, we employ two measures of literacy: literacy scores from the interviewer-administered assessments and a literacy index formed from respondents' self-assessment of literacy abilities and behaviors. For both literacy measures, we find that adults who have lower levels of literacy have a higher item-nonresponse rate across all questions and on questions with or without show cards than those who have higher levels. We then categorize the questions by reading difficulty levels, and find an interaction effect between literacy levels and question difficulty. People with lower literacy levels have substantially higher item nonresponse rates on difficult questions than people with higher literacy levels. This paper concludes with implications of these findings for survey researchers and practitioners and suggestions for future research.

Nexus Between Breakoff and Unit Nonresponse.

Emilia Peytcheva, RTI International (emilia@umich.edu); Andy Peytchev, RTI International (andrey@umich.edu)

Breakoffs in interviewer-administered surveys are different from those in self-administered modes. They are more likely to occur at logical breaks in the survey, arguably because they can be interpreted to some degree as separate survey requests (Groves and Kahn, 1979). Unlike in self-administered surveys where breakoffs occur more uniformly over the course of the survey (e.g., Peytchev, 2009), in interviewer-administered surveys breakoffs occur when perceived separate survey requests are introduced (e.g., between screening and main interview, and at section breaks such as the introduction of new topics and modules). Therefore, it is possible that breakoffs in interviewer-administered surveys share common causes with unit nonresponse. If so, methods to reduce as well as adjust for unit nonresponse may also be effective in addressing bias in face-to-face survey estimates due to breakoff; if not, new methods and separate models will be needed to address breakoff and nonresponse, as has been suggested for self-administered surveys.

We look at common causes for unit nonresponse and breakoff using paradata from the National Health Interview Survey available on complete respondents, breakoffs, and nonrespondents. Paradata come from the Contact History Instrument, which includes summary descriptions of the contact attempt as well as survey organization actions such as leaving a note, "staking out" the household, and checking with neighbors. The paradata also include timing data and keystroke data that can be used to identify common causes, when related to the likelihood of breakoff and likelihood of nonresponse.

We conduct two sets of analyses to address the degree of similarity between breakoff and unit nonresponse; we examine a causal link through common covariates and compare the direction and magnitude of bias in estimates due to likelihood of breakoff and of unit nonresponse. Implications for survey practitioners are discussed.

Verbal Paradata And Income Nonresponse: How Respondents Answer Survey Questions Can Predict Whether And How An Income Questions Will Be Answered

Matt Jans, U.S. Census Bureau (mattjans@isr.umich.edu)

Income nonresponse is a significant problem in survey research, with rates as high as 50% for some types of income. We know little about why it occurs. The way respondents answer survey questions (voice and speech characteristics and question-answering behavior) may predict whether they will provide income data. Audio recordings of interviews conducted by the Reuters/University of Michigan Surveys of Consumers (n=185) were selected based on how respondents answered the household income question. Nonrespondents reported no income information, bracketed respondents reported their income within a range of values, and dollar amount respondents simply provided a dollar value. Respondent utterances were transcribed and coded for speech behavior (e.g., disfluency) and question answering behavior (e.g., qualifications), while voice pitch was extracted mechanically. Listener ratings of psychological states were also applied. These verbal paradata were hypothesized to reflect respondents' affective and cognitive states that make up the psychological mechanisms that may lead to income nonresponse. Statistical results show that verbal paradata can distinguish between income nonrespondents and respondents, even when only using verbal paradata that occur before the income question. This suggests that income nonresponse may be due to problems with survey responding in general. Income nonrespondents have lower affective involvement and express more negativity before the income question than bracketed respondents or dollar amount respondents. Bracketed respondents express more signs of cognitive difficulty than nonrespondents and dollar amount respondents. Further, income nonresponse is best predicted by behavior before the income question, while bracketed income response is best predicted by indicators on the income question itself. The results have implications for theory and research about income nonresponse, specifically the use of verbal respondent paradata to understand the subjective psychological states respondents experience when they answer survey questions, and how those states predict whether income is reported.

Use of Surname and Geocoded Address Information to Target Spanish-Preferring Individuals in a Population Survey with Bilingual Mailings

Marc N Elliott, RAND (elliott@rand.org); Amelia M Haviland, RAND Corporation (haviland@rand.org); David J Klein, Children's Hospital Boston, Harvard Medical School (David.Klein@childrens.harvard.edu); Ron D Hays, University of California, Los Angeles, Department of Medicine (drhays@ucla.edu)

Response rates for Spanish-speaking individuals are often low in large population surveys when budgets do not permit mailing bilingual instruments to all. The Medicare CAHPS Survey, the primary means of assessing the healthcare experiences of American seniors, surveys over 600,000 individuals annually via two mailings and telephone follow-up of an approximately 100-item instrument. Because administrative data lack reliable indicators of language preference, beneficiaries currently receive a bilingual postcard noting the option of requesting a Spanish-language survey. But without an active request that few make, beneficiaries receive only English mailed surveys, and the majority of Spanish-language completes occur during the more expensive telephone follow-up phase.

Immediately sending a double mailing of the survey (English and Spanish versions) to a targeted sub-sample of beneficiaries likely to prefer Spanish might cost-effectively increase overall response rates in this hard-to-reach population and replace some phone interviews with mail completions, but requires the ability to predict Spanish language preference. Using 600,000 records, we modified a surname and address approach previously used to impute race/ethnicity and estimate health disparities in order to predict language preference. Specifically, a logistic regression predicted whether a given beneficiary responded to the survey in Spanish (via mail or phone) from administrative race/ethnicity information, age, surname, a low income indicator, and geocoded information, including linkage to ACS data on the proportion of the population in each county who have both Limited English Proficiency and prefer Spanish. Spanish language preference was predicted with an area under the curve of 97%. We describe this model, which can be used to target double-mailings and allocate bilingual interviewers, and an ongoing experiment, which randomizes high probability individuals to receive double mailing in this year's version of this survey, to estimate the impact of double mailings on response rates (overall and by mode), as well as responses.

Methodological Briefs: The Continuing Challenge of Non-response

Imputing The Missing Y's: Implications For Survey Producers and Survey Users

Rebekah Young, The Pennsylvania State University (rly116@psu.edu); David R. Johnson, The Pennsylvania State University (drj10@psu.edu)

Survey practitioners have an increasing number of tools for handling in an unbiased manner item-level missing data that is a common occurrence in survey datasets. These have increased the utility of planned missing designs which can increase the efficiency of surveys, and have motivated the release of datasets in which all missing values have been imputed to facilitate the analysis of the data by researchers and policy makers. Recent studies, however, have generated serious concerns about the best practices for statistical analysis with an imputed dependent variable (DV). In particular, von Hippel (2007) has argued that the DV should be included in the imputation model but the imputed values on this variable of interest should be removed prior to analysis. He suggests that failure to do so can lead to biased estimates and standard errors. If retaining cases with imputed outcome variables proves to be a serious problem, then survey producers need to release imputed datasets with indicators for all imputed values so that survey users are able to take out the imputed values and delete the cases when the variable is treated as the outcome. Moreover, deleted cases that are missing on the outcome may lead to other biases and prove complex in studies with variables that are outcomes in some analyses and predictors in others. The simulations used by von Hippel may not reflect the complexities of how the practice of imputing then removing the imputed values will fare in analyses of real survey data. In this paper, we compare approaches to handling the imputed DV under several conditions typical of survey data. We also use simulated data to test the performance of this method under realistic outcomes. Our results will help researchers make better informed decisions on how to handle imputed DVs in their analyses.

Nonresponse Followup and Analysis in a Survey of Medical Professionals

Carl M Ramirez, US Government Accountability Office (ramirez@gaio.gov); Shana B Wallace, US Government Accountability Office (wallaces@gaio.gov)

In addition to encouraging response, the followup contacts we made to nonresponding doctors, physician's assistants and nurse practitioners in a 2009 survey brought to light a propensity to nonresponse among likely ineligibles or those of borderline eligibility. Because topic salience is a likely factor in the decision to participate in a survey, we were not surprised to discover that those medical

professionals no longer certified to practice certain occupational health specialties (a characteristic known in the physician assistant sample frame) were less likely to respond to questions concerning issues confronting this field. The discovery of this relationship led to a redefinition of the target population to exclude the subgroup of known non-certified PA's from the eligible population. In addition, analysis of reasons a subsample of nonrespondents gave for not participating were largely eligibility-related conditions such as a focus on other specialties, career change, or retirement, rather than opinion- or experience-specific reasons that might more directly influence levels of bias. Therefore, we conclude that the multiple techniques we used to examine the possibility of nonresponse bias ("two phase" sampling of a subset of nonrespondents for high-effort conversion, administration of a quick, three key question interview to nonrespondents only willing to participate if burden was reduced, and asking why nonrespondents did not want to participate) not only allows examination of nonresponse bias, but reduces overcoverage of ineligibles not otherwise identifiable.

Analysis of Late Responders to Probability-Based Web Panel Recruitment and Panel Surveys.
Yelena Kruse, Knowledge Networks (ykruse@knowledge networks.com); Erlina Hendarwan, Knowledge Networks (ehendarwan@knowledge networks.com); J. Michael Dennis, Knowledge Networks (mdennis@knowledge networks.com); Charles DiSogra, Knowledge Networks (cdisogra@knowledge networks.com)

Transitioning from a ten-year legacy of RDD sampling, the probability-based online KnowledgePanel® uses an address-based sample (ABS) frame for mail-based recruitment of panel members. This paper analyzes nonresponse through a statistical analysis of "late responders" at two stages of the panel survey process: panel recruitment and the online surveys themselves. The characteristics of "late responders" provide a statistical, however imperfect, window into the characteristics of non-responders at these two stages. Furthermore, any differences between early and late responders would have implications for nonresponse bias in the event the field period is shortened for either the panel recruitment or the online surveys. First, to examine late responders to the ABS-sourced panel recruitment, we compared full demographic and other profile-based information that had been collected among both early and late responders to the mail-based recruitment. After these mail-based recruitment efforts, non-responders are contacted using a telephone-based recruitment whenever a landline phone number can be matched to a residential address. Second, to examine late responders to KnowledgePanel web surveys themselves, we analyzed the samples for three large-sample online surveys. Preliminary research found that late responders to the mail recruitment and panel surveys are disproportionately non-white (p-value<0.05), live in three- or more-person households (p-value<0.1), and younger adults (p-value<0.01). The findings will be useful in designing more effective recruitment and online methods encourage response (such as using a targeted segmentation of messages and incentives) in order to minimize response bias.

Cost-Effective and Cooperative Consent: Strategies for Obtaining Parental Permission in Early Childhood Evaluations

Elisha Smith, Mathematica Policy Research (esmith@mathematica-mpr.com); Annalee Kelly, Mathematica Policy Research (akelly@mathematica-mpr.com); Danna Basson, Mathematica Policy Research (dbasson@mathematica-mpr.com)

Approaches for obtaining parental consent in surveys vary greatly. Although there is an extensive literature on the benefits of passive versus active consent, there is less information about what types of strategies are most cost-effective and result in higher rates of participation.

Our methodology brief extends previous work by examining the differences in cooperation and cost-effectiveness between two strategies for obtaining active parental consent: (1) using field enrollment specialists (FES) and (2) using center and program staff.

The Head Start Family and Child Experiences Survey (FACES) was first conducted by Mathematica in 2006 to examine the developmental progress of children and their families during and after Head Start participation. As part of FACES 2009, field enrollment specialists visited Head Start programs up to three weeks before data collection to sample children and classrooms. Then they worked with school staff to distribute and collect consent forms.

The Universal Preschool Child Outcomes Study (UPCOS) was first conducted by Mathematica in 2007. As part of UPCOS, Mathematica collected consent forms from centers and family child care providers through center liaisons and site directors. After the child care providers gathered as many consent forms as they could, field staff conducted in-person follow-up at each site.

This paper will compare the cooperation rate achieved using the FES strategy versus the rate achieved using the center dependent strategy. It will also compare the costs of preparing for fielding and conducting data collection in the two consent approaches. We hypothesize that the FES strategy will result in higher rates of cooperation among teachers and lower costs per complete in the overall study than the center dependent strategy.

Evaluation of an Alternative Nonresponse Adjustment Method

Jared Coopersmith, Education Statistics Services Institute - Quality Information Partners (ctr_jcoopersmith@air.org); Danielle Battle, Education Statistics Services Institute - AIR (dbattle@air.org)

In response to questions about the nonresponse adjustment for the 2004-05 Teacher Follow-up Survey (TFS), sponsored by the National Center for Education Statistics (NCES), which used predefined weighting class cells, the authors tested a new method for assigning the weighting classes using Chi-Square Automatic Interaction Detection (CHAID). CHAID uses chi-square tests of association between characteristics known for the sample and response propensity. These tests of association can then be used to define weighting classes for nonresponse adjustment. The weighting classes produced by CHAID were used to produce two new nonresponse adjustments. The authors compared these two new versions of the nonresponse adjustment using CHAID with the original nonresponse adjustment produced with the predefined weighting classes. The nonresponse-adjusted weight produced with CHAID reduced nonresponse bias more effectively than the original nonresponse adjustment. The CHAID nonresponse weighting cells were also produced more quickly than the original nonresponse weighting cells. The authors concluded that CHAID was the preferred method for determining the weighting classes used for nonresponse adjustment. These findings were accepted and CHAID will be used to produce the nonresponse adjustment factor for the 2008-09 TFS.

In Defense of the Continuum of Resistance: Evidence from a Mail Survey.

Rebecca Medway, Fors Marsh Group, LLC (rmedway@forsmarshgroup.com); Scott Turner, Fors Marsh Group, LLC (sturner@forsmarshgroup.com)

Survey practitioners have adopted a variety of strategies intended to combat falling response rates. For example, in mail surveys, follow-up mailings are commonly used to convince resistant sample members to participate. However, it is important to verify that the responses garnered from follow-up mailings not only increase the response rate but also reduce nonresponse error. The continuum of resistance model provides a framework for considering this issue; it proposes that reluctance to complete survey requests is a meaningful dimension along which sample members vary. The model proposes that (a) late respondents (those who reply to follow-up contacts) differ from early respondents (those who reply to the

initial survey request), and (b) late respondents serve as a good proxy for nonrespondents. However, the model has experienced mixed success when applied to empirical data. This study contributes to the debate over the applicability of the model.

In order to test the continuum of resistance model, responses from a 2008 mail survey of military recruiters were paired with administrative records that provided information about the characteristics of nonrespondents. Analysis of this data supported both tenets of the continuum of resistance model. Although the characteristics of early and late respondents were relatively similar, the two groups provided significantly different responses to half of the twelve survey questions that were examined. In general, early respondents' attitudes were more extreme and negative than those of late respondents. Late respondents were also found to be potentially a good proxy for nonrespondents, as they displayed fewer demographic differences from nonrespondents than did early respondents. The few characteristics that did vary between late respondents and nonrespondents did not significantly affect responses to survey questions.

Data Quality and the Reluctant Respondent: Evidence from the Spanish Survey of Household Finances.

Sarah M Hughes, NORC (hughes-sarah@norc.org)

Survey sponsors focus much time and expense on improving response rates, and methodological researchers focus much time analyzing whether it is worth the effort in terms of reducing nonresponse bias. This methodological brief presents evidence from the Bank of Spain's Survey of Household Finances on the quality of data gathered from reluctant respondents as compared to that gathered from cooperative respondents. The analysis rests on specific measures of data quality including item nonresponse, timing data, exact values vs. range values for monetary response data and other quality measures. Reluctant or cooperative respondents are defined by refusal behavior evident in the paradata gathered during fieldwork. The analysis is expected to contribute to the dialogue surrounding nonresponse in comparative survey research.

Methodological Issues with Address Based Sampling

Comparing the Availability & Accuracy of Sample Frame Indicators Across Two Sampling Frames: Address Based Sampling vs. Telephone

Anh Thu Burks, Nielsen Company (anh.thu.burks@nielsen.com); Michael W Link, Nielsen Company (michael.link@nielsen.com); L. Tracie Yancey, Nielsen Company (tracie.yancey@nielsen.com); Natalie Coser, Nielsen Company (natalie.coser@nielsen.com)

Address based sampling (ABS), when compared to random digit dialing (RDD) landline telephone frame, has become a viable alternative as a sampling methodology, especially given the issues commonly associated with the use of RDD (i.e., exclusion of cell phone only homes, issues related to number portability, and decline in participation of younger hard-to-reach demographic). Both the address and telephone frame offer the potential to append a vast array of demographic information from sample vendors (i.e., age indicator). This information can be used by researchers to tailor specific types of mailings and incentives toward respondents in an attempt to gain their cooperation. With ABS being a relatively new sampling methodology, limited focus has been placed on the types of data that are readily available in ABS and how they may differ in accuracy to data that are available in RDD. Furthermore, comparisons have not been made to assess the overall accuracy and quality of these types of indicators across ABS and RDD. Thus, the focus of this paper is to juxtapose ABS and RDD and to examine three

areas of interests: (1) the types of information that is innate and specific to each frame, (2) the availability and accuracy of demographic indicators (age, Hispanic surname, etc.), and (3) the focus on the areas of improvement for the utility of these indicators. More specifically, we will compare the February 2009 ABS-based measurement with the February 2008 RDD-based measurement taken from Nielsen TV Ratings Diary Survey. The differences and findings noted between these sampling approaches will be discussed within our broader understanding of the benefits that each frame has to offer and how these indicators can be used to enhance the survey recruitment process.

From Mailbox to Computer: The Effects of Priority Mail & a Second Incentive on Internet Response using Address-based Sampling in a General Public Survey.

Benjamin L. Messer, Washington State University Social and Economic Sciences Research Cent (aluvial@gmail.com); Don A. Dillman, Washington State University (dillman@wsu.edu)

It has been proposed that the near complete coverage of residential postal addresses in the U.S. Postal Service's (USPS) Delivery Sequence File (DSF) provides a more comprehensive sample frame for general household surveys than random digit dialing. In addition to obtaining responses by mail, it may be possible to convince many householders to submit survey responses over the Internet. However, effective procedures for achieving that goal are not yet well developed. In this experiment we explore the potential for using different postal contact strategies with DSF households in order to obtain an Internet response. It builds upon previous research by the authors which showed that response rates of 45% could be obtained in statewide sample surveys of 20 minute questionnaires using token financial incentives and certain other design features, with 2/3 of those responses from the Internet and the remainder by postal mail. A new experiment, conducted in October-November 2009, seeks to increase the number and proportion of households responding over the Internet (as opposed to mail) through the use of a USPS Priority Mail follow-up contact and an additional token cash incentive in the follow-up contact. We also compare answers to those received by a mail-only approach and examine non-response error through comparisons with recent American Community Survey results for the state of Washington. The overall goal of our research is to develop a viable data collection method utilizing the DSF and the Internet to conduct general household surveys as a possible alternative to random digit dialing.

A Follow-Up to Address Based Sampling and Address Matching: Experience from REACH U.S.

Katie Dekker, NORC (dekker-katie@norc.org); Whitney Murphy, NORC (murphy-whitney@norc.org); Kanru Xia, NORC (xia-kanru@norc.org)

Address-based sampling (ABS) is becoming a more widely used alternative to random-digit dialing. A multi-mode approach using ABS typically includes a telephone component for addresses that are matched to phone numbers, a mail component for those that are not matched, and sometimes web and face-to-face components. Because telephone interviewing is still one of the most efficient and effective modes of data collection, a successful address-based approach relies heavily on accurate matching of addresses to working telephone numbers.

Using this design, NORC, under the sponsorship of the Centers for Disease Control and Prevention, began conducting Phase 1 of the Racial and Ethnic Approaches to Community Health Across the U.S. Risk Factor Survey (REACH U.S.) in 28 communities. Over 24,000 adults were surveyed by a combination of phone, mail, and in-person interviews.

After selecting a sample of addresses, NORC matched those addresses to telephone numbers using two vendors. We expected more than half of the addresses to be matched to a working phone number, and thus to be interviewed via telephone. During the telephone interview, respondents were asked to verify

that they lived at the selected address. If the address was not verified, indicating that the matching process did not successfully link the household to an appropriate telephone number, the sampled household was surveyed via another data collection method.

The matching success rate was lower than expected, and the matches were less accurate than desired, especially in particular communities and housing types. Details of the sampling approach, along with some Phase 1 results, will be presented to show the wide range of matching rates and its impact on data collection. Decisions for future rounds of data collection that were made based on these results will also be discussed.

Under the Hood of ABS: An Analysis of Call Outcomes, Call Attempts and Contact Times
Kimberly Brown Hawkins, The Nielsen Company (kimberly.hawkins@nielsen.com); Michael A Howton, The Nielsen Company (Michael.Howton@nielsen.com); Anh Thu Burks, The Nielsen Company (Anh.Thu.Burks@nielsen.com); E. Nicole Bensky, The Nielsen Company (Nicole.Bensky@nielsen.com)

Survey researchers who have converted to an Addressed-Based Sampling methodology may wonder how calling patterns have changed in the new ABS environment and the magnitude of these changes in comparison to the RDD environment. Compared to RDD sampling, ABS sampling begins with addresses for each record, increasing frame coverage by providing the possibility of a mail contact method if there is no phone number match. Because the ABS methodology increases coverage to 98% from around 70% in RDD, of interest will be seeing how previously under-sampled households affect overall calling patterns. Matched sample, where addresses are matched to a phone number and treated similarly to listed RDD sample, now includes slightly more households comprised of hard-to-reach demographics in comparison to RDD. Additionally, unmatched sample (addresses that do not have a matching phone number are initially contacted via a mail survey and then contacted by phone if the household provides a phone number) now includes cell-phone only homes and unlisted homes in zero listed banks, which also have the potential to affect calling patterns. In 2008, The Nielsen Company migrated to an Addressed-Based Sampling frame for the TV Diary service, and currently phones these two types of ABS sample: matched sample and unmatched sample. This paper will compare previous RDD calling pattern analyses with ABS calling pattern analyses and present differences in contacts and completes by call attempt, final call outcomes and best contact and completion rates by time of day and day of week. Additionally, differences in call history behavior for landline and cell phone records will be presented for unmatched sample. Survey researchers interested in ABS methodologies will find these results relevant for staffing, budgeting, sample representation, response rates and calling rule purposes.

Designing Multi-Mode Address-Based Methodologies to Target Specific Ethnic Communities: The Case of REACH U.S.
Edward M. English, NORC (english-ned@norc.org); Colm O'Muircheartaigh, NORC (colm@norc.org); Kanru Xia, NORC (xia-kanru@norc.org)

The Centers for Disease Control and Prevention (CDC) has awarded NORC a five-year contract to conduct the Racial and Ethnic Approaches to Community Health Across the U.S. (REACH U.S.). REACH U.S. is the cornerstone of CDC's efforts to eliminate racial and ethnic health disparities in health priority areas including breast and cervical cancer, cardiovascular disease, diabetes mellitus, adult/older adult immunization, and hepatitis B. In 2007, CDC established cooperative agreements with 40 organizations for each to develop and implement innovative strategies to reduce health disparities among specific ethnic and racial minority populations. These populations include African Americans, Native Americans, Hispanic Americans, Asian Americans, and Pacific Islanders. NORC is in the process of conducting Risk

Factor Surveys (RFS) in 28 REACH U.S. communities, and the RFS will provide CDC and the grantees with quantitative data to track the progress and achievements of grantee communities.

REACH U.S. has employed an address-based sampling approach (ABS) combined with data collection via telephone, mail, and face-to-face interviews to obtain the requisite number of completes in each community. The 28 communities were quite different in a number of key aspects, including target population, urbanicity/rurality, eligibility rates, population density, language requirements, and literacy. As such, the outcomes for the various communities were often dissimilar. We compare our assumptions of eligibility and productivity to our experience across mode and across communities. Our research provides guidelines for how to conduct multi-mode ABS studies in targeted race/ethnic communities in a variety of circumstances.

Pre-election & Election Polling

An Attempt to Substitute Cell Phone Interviews with Non-Cell Phone Interviews by Sample Matching

Masahiko Aida, Greenberg Quinlan Rosner Research (maida@gqrr.com)

The goal of this research is to answer the following question, “is there an effective and systematic way to substitute interviews from cell phone sample frame with alternative (cheaper) sample frame?”

While research examining effect of including cell phone interviews on 2008 presidential election polls showed only minor differences (Keeter et al, 2008), this finding is not guaranteed to hold true in the future. Therefore, pollsters have two options, the first is to secure funding to conduct cell phone interview supplements, the second is to forget about the cell phone problem but live in a fear of under coverage and skepticism for his own data. Furthermore, there are situations when conducting cell phone supplements are not feasible (ex. polling on congressional level races) for technical issues concerning cell phone exchanges.

This research will offer an alternative option to pollsters, an analytical tool that allows them to find “cell phone respondents look likes” from interviews from other (and cheaper) sample frames. This tool is called sample matching, originally a means to evaluate causal inferences from observational studies when rigorous random assignment is absent. Douglas Rivers and his company (YouGov Polimetrix) pioneered in applying matching idea in order to draw “representative” sample from opt-in panel.

We will employ same idea in order to draw “cell phone look like sample” from pool of landline respondents using propensity models created by machine learning algorithms.

The research demonstrates how pollsters can analytically synthesize cell phone like sample using sample matching. We will also present situations when bias reduction due to matching fail due to model miss-specification of the propensity models.

Online Polls and Registration Based Sampling: A New Method for Pre-election Polling.

J. Quin Monson, Brigham Young University (Quin.Monson@byu.edu); Kelly D. Patterson, Brigham Young University (Kelly_Patterson@byu.edu); Chris Mann, University of Miami (cmann@miami.edu); Michael Barber, Princeton University (barber@princeton.edu)

This paper presents a new method for pre-election surveys by combining the best features of other pre-election survey methods. This survey is administered online to a probability sample drawn from a voter registration list. Respondents are randomly selected from the voter file and sent an invitation letter with a link and an access code for an online poll. Two methods of probability sampling are used: a Simple Random Sample (SRS) and a Probability Proportionate to Size (PPS) sample. The PPS sample is drawn using a regression model employing variables available in the voter file to produce a probability of voting for each individual in the voter file. The paper examines data from four elections that cover a variety of electoral situations including a low turnout primary, a low turnout municipal general election, and a high turnout statewide election. Three of the elections are in Utah and the fourth is in Florida. Preliminary results indicate that the methodology is very accurate compared to election results. Furthermore, parallel exit polling allows for a closer examination of the composition and accuracy of the pre-election poll samples with that of available exit poll data. When the SRS and PPS sampling strategies are compared, the PPS sample is especially accurate in the low turnout elections where producing an accurate likely electorate has historically proven to be difficult. A counterintuitive possibility of this research is that unlike most surveys where lower response rates increase the chance of non-response error, the low response rates in these surveys may actually decrease the chance of non-response error. Low response rates could improve the coverage of the population of likely voters because the time and effort to go online and complete the questionnaire mimics the act of voting and may require the similar levels of political interest,

Understanding the Role of Participation Rate Changes and Preference Changes in Electoral Outcomes: Results of the Analysis of Homogenous Sub-Group Time Series Data

Chris Wilson, Wilson Research Strategies (cwilson@w-r-s.com); Bryon Allen, Wilson Research Strategies (ballen@w-r-s.com)

Analyses of the 2008 and 2009 elections have focused on two dueling narratives:

- Narrative 1: The difference between the 2008 and 2009 results were primarily a result of differences in participation rates of key groups—in particular young voters and minorities.
- Narrative 2: The difference between the 2008 and 2009 results were primarily a result of difference in preferences among so-called “swing voter” groups.

If the first narrative is a “participation hypothesis” while the second is a “preference hypothesis,” understanding which of these hypotheses is a better descriptor of actual electoral behavior methodological and analytical discussions regarding weighting of voter samples and analysis of survey-to-survey variations in intended vote and other metrics.

While several analyses have explored the participation and preference hypotheses in the context of the 2008/2009 elections, we analyze a larger time-series of data to attempt to evaluate which these hypotheses explain electoral behavior. We combine National Election Study dating back to 1980 and data from exit polls and post-election studies for “off-year” state elections to arrive at a cumulative time-series data set covering more than 20 years and several shifts in political control.

We construct a series of homogenous voter sub-groups from the combined data set using latent class analysis. For each homogenous sub-group, we estimate a participation trend function and a preference function. These functions were then tested as potential explanatory variables for observed outcomes.

Preliminary results suggest that for most sub-groups in the analysis, participation rates are more important than preference changes in explaining election-to-election variations in outcome. For a few sub-groups of the electorate, however, preferences are highly volatile and for these sub-groups the

preference hypothesis strongly holds. For some election-to-election changes, the preference shifts of these few volatile groups explain much or all of the change in observed outcomes.

***Confirming the Validity of Survey Respondent Reports of Voter Registration and Turnout:
Checking the Records Turns Up Surprising Bad News***

Matthew Berent, Stanford University (mberent@stanford.edu); Jon A Krosnick, Stanford University (krosnick@stanford.edu); Matthew DeBell, Stanford University (debell@stanford.edu)

Many social scientists presume that government records can be used to accurately measure whether survey respondents are registered to vote and whether they did in fact vote in a particular election. In fact, some researchers believe that surveys need not ask respondents whether they are registered or voted - simply getting respondents' names and addresses (and perhaps birth date) is sufficient allow researchers to reliably determine whether each respondent was registered and voted. In this paper, we evaluate the validity of such assumptions. Using data from the American National Election Studies 2008 surveys, we compared survey reports to government records of turnout. Surprisingly, we found that it is essentially impossible to use government records to confidently determine whether a survey respondent was registered and/or voted. The government files from multiple states clearly contain errors, so the process of identifying a government record for each survey respondent (or determining that there is no such government record) is impossible to do with certainty. And depending upon the particular decisions made by researcher in using the records, very different conclusions can be reached about how many respondents were registered and how many voted. All this casts serious doubt on the validity of using government records instead of self-reports of registration and turnout.

AAPOR Concurrent Session A

Saturday, May 15, 2010

8:00 a.m. - 9:30 a.m.

AAPOR Task Force Report on Online Panels

The AAPOR Online Panel Task Force

Reg Baker, Market Strategies International (reg.baker@marketstrategies.com); Members of the AAPOR Online Panel Task Force

In 2008 the AAPOR Council established an Online Panel Task Force with the charge to review the current empirical findings relative to online panels used for data collection and to make recommendations to AAPOR members regarding their use. The work of the Task Force is now complete. In this session members will describe the overall approach taken and the reasoning behind the report's conclusions and recommendations. The session is meant to provide an opportunity for AAPOR members to interact with the Task Force, to raise issues important to them, and in general to understand the report's conclusions and recommendations relative to this major new sampling and data collection method.

Address Based Sampling Issues & Challenges

Findings from a Two-Phase Mail Survey for a Study of Veterans

Daifeng Han, Westat (daifenghan@westat.com); David Cantor, Westat (DavidCantor@westat.com); Pat Dean Brick, Westat (PatDeanBrick@westat.com); Maribel Aponte, Department of Veterans Affairs (Maribel.Aponte@va.gov)

Address-based sampling has been used in response to the declining coverage and response rates of random digit dial surveys. Although mail surveys using address-based samples have proven to be a promising approach in general population studies, their usability for studying specific subpopulations is untested. Since there is no interviewer to screen households for the targeted group, it is not clear how well a passive method like a mail survey would work. This paper reports findings from a two-phase pilot study surveying Veterans as part of the National Survey of Veterans (NSV) sponsored by the US Department of Veterans Affairs. The Pilot demonstrated the feasibility of applying two-phase mail survey design for Veterans and distinguished some factors associated with response rates.

The first phase of the pilot was a mail screener sent to a national sample of 11,000 residential addresses selected from the USPS address lists. The second phase was a topical questionnaire directed to the Veteran identified through the screening. The screener inquired about the existence of veterans in the household as well as their demographics and yielded an overall response rate of 35% among all households. Using the data from the Department of Veteran Affairs, we estimated the total number of veterans in the US and found that the screener respondents accounted for 66% of all veterans. The topical questionnaire was administered to all the veterans identified through the screening and achieved a response rate of 60%.

The Pilot Study included stratified sampling to oversample addresses with a higher prevalence of Veterans and several experiments. This paper will focus on the: (1) response rates and effective coverage rates by demographic characteristics; (2) enclosing a colored information page in the screener package targeted to improve response; and (3) stratified sampling to increase the efficiency of the screening.

Capturing Change in the USPS Delivery Sequence File: A Spatial-Temporal Analysis.

Colm O'Muircheartaigh, NORC at the University of Chicago (O'Muircheartaigh-Colm@norc.org); Ned English, NORC at the University of Chicago (english-ned@norc.org); Michael Latterner, NORC at the University of Chicago (latterner-michael@norc.org)

Address-based sampling (ABS) has emerged in recent years to remedy the growing coverage problems with RDD sampling. The United States Postal Service's Delivery Sequence File (DSF) has made ABS possible and cost-effective. Research conducted by survey firms has validated capacities of the DSF through in-person verification of the DSF in a cross-sectional set of areas. Such research has also revealed artifacts of an improving DSF. The present research is focused on exploring this changing coverage by analyzing the DSF over the past several years on both a macro level and a micro level.

First, we present yearly estimates of the percentage of units city-style for the entire U.S. with measures of change over time. Percentage of units city-style has been a main indicator of coverage and understanding how this is changing on a national level will inform researchers as to what we should expect from the DSF in the coming years. Second, we examine how previous estimates of coverage can help predict coverage on a micro level by looking at areas for which we have verified coverage in past years and explore the change to the DSF before and after our verification. Finally, we describe the various data sources that together may explain change in DSF coverage (i.e. housing start data, 911 conversion, change in mail delivery).

A Comparative Study of ABS and RDD Sampling in Two Locations: The Commonwealth of Massachusetts and The District of Columbia.

Susan Sherr, SSRS (ssherr@ssrs.com); Timothy Triplett, Urban Institute (TTriplett@urban.org); Robyn Rapoport, SSRS (rrapoport@ssrs.com); Sharon Long, Urban Institute (slong@urban.org); David Dutwin, SSRS (ddutwin@ssrs.com)

The paper compares the results of a Random Digit Dial (RDD) sample design with a United States Postal Service (USPS) address list or ABS sample design in two locations: The Commonwealth of Massachusetts and the District of Columbia. The Urban Institute and Social Science Research Solutions collaborated in conducting the evaluation. While the overall plan in both studies was to combine the two samples, a secondary goal was to choose large enough samples of both the RDD and the USPS so that they could produce independent estimates as well. Both studies evaluate programs to maximize health insurance coverage for citizens. The authors have previously presented results from the Massachusetts study demonstrating that the ABS sample produced younger respondents and those from cell-phone-only households, thus providing evidence that ABS helps to address two critical coverage problems associated with RDD samples. Here, they will compare these favorable results from Massachusetts with the outcome of a more recent, identical design in a racially-diverse urban area with a higher concentration of low-income residents. This is of particular interest because results from Massachusetts indicate that an ABS design that requires unlisted respondents to complete a questionnaire on the web or through the mail (or to call in to a call center) does not increase the percent of low-income respondents in the sample. The research provides insight into the coverage and cost/benefit trade-offs that researchers can expect from RDD sample designs versus USPS address sample designs that utilize a multi-mode (phone, web,

mail) data collection approach and evaluates whether these results differ between a state-wide survey and one in an urban area.

Applying Address-Based Sampling as an Alternative to RDD in a Community Survey

Gene M. Lutz, University of Northern Iowa (gene.lutz@uni.edu); Mary E. Losch, University of Northern Iowa (mary.losch@uni.edu); Disa L. Cornish, University of Northern Iowa (disa.cornish@uni.edu); Anne Bonsall Hoekstra, University of Northern Iowa (anne.bonsallhoekstra@uni.edu)

Address-based sampling (ABS) has been proposed as a promising alternative to random digit dial (RDD) surveys; the sampling method may ease the challenge posed by increasing coverage bias in RDD samples due to the growing number of cell-only households. To explore the utility of the method, ABS was used to conduct a mixed mode (web and telephone) survey to measure public opinion of city services among residents of a small Midwestern city. Informational letters were sent to 3,000 households inviting eligible residents (the adult with the most recent birthday) to participate in the survey via one of two modes: online or telephone. Residents were provided a contact information card to return with their first name, preferred telephone number, and preferred time to be reached if they desired phone contact. Available phone numbers were dialed for those not returning a contact card or responding by web. Consistent with estimates of telephone number availability in ABS sample frames, in the original sample of 3,000 addresses, 57% had a telephone number attached to the address and 87% were valid addresses (not returned as undeliverable). Data collection through combined modes yielded 836 completed interviews: 350 online and 486 via telephone. The response rate (AAPOR RR3) was 35% with a cooperation rate (AAPOR CR2) of 88%. Findings regarding the reduction in coverage bias were mixed. Although coverage was improved by inclusion of a significant number of cell-only households (13% of the sample), the present study did not reach a greater proportion of young adults than an earlier survey of the same community. Additional findings will be presented regarding behavior of the sample, comparisons with other recent local RDD samples, and with reports of other ABS results.

Measuring the Undercoverage of DSF vs. Traditional Frame Construction Approaches: Comparing Potential Address Sources for the National Children's Study

Colm O'Muircheartaigh, NORC (colm@norc.org); Edward M. English, NORC (english-ned@norc.org); Katie Dekker, NORC (dekker-katie@norc.org); Stephanie Eckman, NORC (steph.eckman@gmail.com)

Motivated by the high costs associated with traditional listing, survey research organizations have been undertaking recent research into using the United States Postal Service Delivery Sequence File (DSF) as a replacement for traditional listing, at least in urban areas. At question is how do different versions of the DSF compare with respect to coverage, including files provided by separate vendors or from different points in time.

Working with the National Children's Study (NCS), we continue our work to understand the coverage properties of these databases in order to determine what households tend to be covered vs. omitted. In-field listings for the NCS present the opportunity to compare the coverage of a number of address database sources with a traditionally-listed frame. After matching the listed addresses with three versions of the DSF, we returned to those housing units that were missed by one or more sources to collect additional data at the housing unit and block-face level. We also used commercially-available household-level data to describe qualities of households that were missed by any particular source.

Data Quality in Cell Phone Survey

Investigating Data Quality in Cell Phone Surveying

Paul J Lavrakas, Independent Consultant (pjlavrak@optonline.net); Trevor N Tompson, The Associated Press (ttompson@ap.org); Robert G Benford, GfK Custom Research NA (Robert.Benford@gfk.com)

Although much of the R&D that has been conducted to date concerning cell phone surveys has focused on the issues of coverage, sampling, nonresponse, and weighting, not much has been reported about the quality of the data that are gathered from respondents interviewed on their cell phone. An AAPOR Task Force on Cell Phone Surveying (Lavrakas, et al. 2008) suggested that “there currently is no reliable evidence that the data gathered in good cell phone surveys are of lower quality than in comparable landline surveys. However, the Task Force believes it is advisable that researchers remain attentive to this concern. Future research is needed to know with confidence if, and how, data quality is affected by gathering it from a respondent on a cell phone” (p. 5). In the spirit of this call for further research into cell phone data quality, the Associated Press and GfK Roper have conducted several dual frame national telephone surveys in which approximately 200 adults were sampled and interviewed via their cell phone in each survey. Each cell phone respondent was asked whether s/he was being interviewed at her/his home or someplace else. A hypothesis suggested in the literature is that for many reasons those interviewed at home on their cell phone may, as a group, provide better quality data than those reached on a cell phone someplace else. Preliminary findings from our analyses of four surveys completed to date show that (1) about one-third of cell phone respondents complete their interview away from home and (2) there is more missing data, including more refusals to sensitive questions, among those interviewed away from home when controlling for the demographic correlates of location of the interview.

Threats to Data Quality in Telephone Surveys: Hands-free Headsets, Poor Audio Quality, and Respondent Distractions.

Courtney Kennedy, The Everett Group (ckkenne@umich.edu); Stephen E Everett, Everett Group (see@everettgroup.com)

In the early days of telephone surveys, respondents were interviewed on fixed, residential landlines. Today they can be reached away from home or within the home, moving about from room to room, using a cordless landline. In addition, hands-free headsets afford respondents substantial physical freedom to do other activities while being interviewed. The implications for data quality, specifically measurement error, are not entirely understood. Presumably, being in a distracting environment, experiencing difficulty hearing the interviewer, or doing intensive multitasking increases the probability of misreporting. This study empirically assesses the impact of four factors on data quality: using a hands-free headset, the audio fidelity of the call, being away from home, and engaging in other activities during the interview. The data come from a national two-wave dual frame (landline and cell) RDD survey. The independent variables were measured using respondent self-reports about their activities during the interview and the sound clarity of the call. Data quality is measured using seven indicators of cognitive shortcutting, such as estimating answers to behavioral frequency questions, recency effects, non-differentiation to battery-format questions, length of open-ended responses, and failing to account for important in/exclusions. Understanding the effects on data quality from audio fidelity, respondent mobility, and distractions may help researchers to determine if and when reporting error is more problematic in cell phone surveys than landline surveys. A previous presentation of data from this study investigated the difference between landline and cell phone respondents on the seven cognitive shortcutting indicators. A small number of significant differences were observed. This paper presents a different set of results that look at whether

headsets, poor audio quality, and multitasking explain the error. It extends the previous work by addressing the critical follow-up question of why device differences in data quality exist when we do observe them.

An Experiment Testing the Impact of Leaving Voice Messages in Cell Phone Surveying
Robert Benford, GfK Custom Research North America (robert.benford@gfk.com); Paul J. Lavrakas, Independent Consultant (pjlavrak@optonline.net); Trevor Thompson, Associated Press (ttompson@ap.org); Christopher Fleury, GfK Roper Public Affairs & Media (Christopher.Fleury@gfk.com)

A robust body of literature exists regarding the role of answering machines or voicemail in landline telephone surveys. However, we are aware of no research on this topic among cell phone sample surveys. Whereas landline mailboxes can be individualized for members of a household they generally are generically addressing the telephone number or household reached. It is also the case that not all households have voice mail. On the other hand, cell phones are primarily individual and almost always have at least the capability to set up a voice mailbox. All cell phones have a Caller I.D., that is, the inbound telephone number shows up on a screen built into the telephone. Thus, our working hypothesis is that a voice message on the initial call can alert the potential respondent to the nature of the call and the sponsor behind the number that appears on Caller I.D. This in turn is reasoned to increase the likelihood of eventually completing an interview on a subsequent call. We also evaluate any difference in costs due to this step in the methodology. This paper presents the findings from an experiment across two Associated Press polls that use a dual frame cell and landline methodology.

Incentives for Cell Phone Users: What Difference Do They Make?
Robert W. Oldendick, University of South Carolina (oldendick-bob@sc.edu); Dennis N. Lambries, University of South Carolina (dennis-lambries@sc.edu)

In the past decade, as the percentage of the U.S. population living in cell-phone only households has steadily increased, telephone survey researchers have attempted various methods to incorporate cell phone exchanges into their general population samples. Many of the early studies that incorporated cell phone exchanges included the payment of an incentive to those who completed an interview on a cell phone; part of the reasoning for this was to defray the cost of the cell phone minutes that respondents were using for the interview. As cell phone plans have evolved and more people have calling plans that have unlimited or virtually unlimited minutes, the relevance of “paying for” used minutes has decreased. Given this, the question is whether providing incentives to cell phone respondents continues to make a difference in response rates. Two experiments were conducted in which half of the cell phone respondents were provided a \$10 gift card incentive and the other half were not. Comparisons of the response rates, the demographic composition of the resulting samples, and the responses to substantive items indicate that providing incentives to cell phone participants may not be cost effective. While not providing incentives will not eliminate the cost disparity between landline and cell phone completions, it will help to reduce these differences.

Adjusting for Cell Phone Nonresponse in Dual Frame Samples
Randal ZuWallack, ICF Macro (randal.zuwallack@macrointernational.com); Frederica Conrey, ICF Macro (frederica.conrey@macrointernational.com)

Most cell phone samples result in an overrepresentation of cell-only respondents. This is likely due to differences in contact rates and cooperation rates between various types of cell phone users. Cell phone reliance can be represented as a continuum ranging from those who use their cell phone exclusively (cell-

only) to those who do not use a cell phone (landline only). In between are people who vary on their cell phone reliance--some use their cell phone considerably, while others very little. Since we are contacting respondents through their cell phones it seems intuitive that those who rely more on their cell phone for communication will be easier to reach.

The bias toward those who rely heavily on their cell phones causes dual frame survey estimates to be skewed. One method to adjust for the bias is to weight the phone user groups to known population totals. This method is limited by the availability of current and accurate data for the phone user groups. As an alternative that is not limited by the availability of external data, we have developed a propensity score adjustment. In this application, the propensity score is the conditional probability of being observed in a cell phone survey given known demographic or lifestyle conditions. The propensity scores reflect each subject's propensity to pick up a cell phone and respond to a cell phone survey. The model is built with data from a national cell phone sample and the National Health Interview Survey. The model is applied to a state cell phone survey and a city cell phone survey, where the phone user populations are unknown. In both cases, the model successfully reduces the bias in relation to phone group status allowing for accurate dual frame estimation.

Methodological Briefs: Longitudinal Surveys, Internet Access & Data Quality

Marge! It's been 30 years! Finding Participants for a 50-year High School Follow-up Study
Pia Peltola, American Institutes for Research (ppeltola@air.org); Lynette Zickl, American Institutes for Research (lzickl@air.org); Susan Lapham, American Institutes for Research (slapham@air.org); Cameron McPhee, American Institutes for Research (cmphee@air.org)

This presentation details tracking procedures used in efforts to find current contact information for participants in a longitudinal study who were last contacted more than 30 years ago and evaluates the success of the various tracking methods used. Locating these same individuals after a 30-year hiatus is crucial to determine the feasibility of a 50-year follow-up study to examine the relationship between cognitive and noncognitive skills and life course outcomes. The study, Project TALENT, is a large, nationally representative longitudinal study of men and women who were in high school in 1960. The American Institutes for Research (AIR) collected extensive information in four half days of testing on the characteristics and cognitive abilities of approximately 440,000 high school students. Follow-up surveys were conducted 1, 5, and 11 years after high school graduation. Also, between 1960 and 1978 students were sent an annual newsletter that included a request for contact information updates. However, 1978 was the last contact made with the participants. Starting in 2009, we have employed multiple methods and sources to track individuals who are now in their 60s. The tracking methods we will discuss include information from the Social Security Administration's Death Master File and LexisNexis. We will also compare death and contact information purchased from MicroBilt, PeopleFinders, and LexisNexis and examine the results by gender. Finally we will discuss our use of social networking media for finding Project TALENT participants. These include Facebook, the Project TALENT website, and 50-year reunions. For each method, we will describe how we used it, how successful we were, and the cost. Preliminary results show that 10 percent of the participants have died and current address or phone information has been updated for 77 percent of those who are still living.

What Survey Design Issues are Impacted by Incentive Conditioning?
John M. Finamore, U.S. Census Bureau (john.m.finamore@census.gov)

The National Survey of College Graduates (NSCG) is a longitudinal biennial survey that collects information on employment, educational, and demographic characteristics of the college-educated science and engineering workforce in United States. As a measure to address increasing nonresponse, we experimented with a monetary incentive offer in the 2006 survey cycle to historically low responding groups. Through a designed experiment, the NSCG offered another monetary incentive in the 2008 survey cycle to investigate the presence of an incentive conditioning effect in survey response rate on these panel respondents. Initial findings indicated the use of incentives in the 2006 survey cycle did not create an expectation of an incentive for response in the 2008 survey cycle. In other words, no significant conditioning effect was seen on subsequent survey response rates.

The goal of this paper is to evaluate the impact of incentive conditioning on numerous other survey design issues including data quality, response time, data collection mode, response propensity patterns, and survey estimation. To understand the impact of incentive conditioning on the total survey design, these issues need to be considered in combination with the response rate findings. This paper will present results related to these other survey design issues from the 2006 and 2008 NSCG incentive conditioning experiment.

The Priming Directive: Panel Effects in Media Awareness Measures on a National Probability Panel.

Kara A Marsh, Fors Marsh Group (kmarsh@forsmarshgroup.com); Robert P Daves, The Everett Group (rob@everettgroup.com); Amanda Anderson, Fors Marsh Group (aanderson@forsmarshgroup.com); Scott Turner, Fors Marsh Group (sturner@forsmarshgroup.com); Stephen E Everett, The Everett Group (see@everettgroup.com); H. Allen White, The Everett Group (allen@everettgroup.com)

Internet panels derived from probability-based sampling are attractive to researchers confronting the limitations of using more traditional modes, such as RDD telephone, to obtain nationally representative estimates (particularly of youth). These panels offer superior coverage by using multiple recruiting strategies (e.g., telephone and addressed-based), and they also have the capacity to include both visual and audio stimuli. However, panel conditioning and panel attrition are potential weaknesses that can compromise panel data quality.

New complexities in panel conditioning have arisen from the growing popularity of a relatively scarce resource -- high quality Internet-panels. In this paper, we discuss one such issue where panelists' awareness of advertising stimuli in Study B was inadvertently primed by their exposure to the same or similar stimuli during earlier participation in Study A, where Studies A and B were operating independently -- of each other and with separate clients.

Preliminary results showed that across multiple administrations the proportion of overlap was as high as 34% for concurrently conducted studies, and cumulatively more than 50% of the participants in Study B had also participated in Study A. Preliminary results also revealed that panelists in Study B were twice as likely to recall advertisements they were previously exposed to in Study A.

This paper will provide a more thorough analysis of the data and identify specific repercussions in data quality for both Study A and Study B. It will also provide a discussion of the steps that were taken to work with the vendor to prevent future conflicts. It will close with recommendations not only for future research on panel conditioning, but also for researchers who employ Internet-panels for their own efforts.

Does Providing Internet Access to Non-Internet Households Affect Reported Media Behavior for Latinos and Non-Latinos? Results from a Six-Month Longitudinal Survey

J. Michael Dennis, Knowledge Networks (mdennis@knowledge networks.com); Charles DiSogra, Knowledge Networks (cdisogra@knowledge networks.com)

A distinguishing feature of KnowledgePanel® and KnowledgePanel Latino is the inclusion of households that do not have Internet access at the time of recruitment. KN provides these households with laptops, and in the past, an MSN-TV® device, and free monthly Internet access so they can participate in online KnowledgePanel® surveys. . Over half of the KnowledgePanel Latino households have been provided a laptop.

Does providing internet access to respondents have an effect on their survey responses? Are the respondents “changed” as a result?

To answer this question, KN launched a longitudinal study of recently recruited KnowledgePanel Latino members measuring any behavioral changes occurring over time with regard to Internet usage, media usage, online shopping, and also measuring comfort with technology. All study subjects were surveyed at 3-month intervals. This presentation will show findings based on three measurement waves for the Internet and non-Internet households. Statistical evidence supports the hypothesis of a limited “novelty effect,” that is, panelists on a limited basis increase their usage of some media and visits of websites when first joining the panel, but then return to lower usage levels by month six of their panel tenure.

Discrepancies Between Student Reports of Parental Occupation and Level of Education – Do They Really Matter? Evaluating Self-Reported Socioeconomic Status Using NLSF Data, 1999-2003
Melissa Bert, Harvard Graduate School of Education (mrb832@mail.harvard.edu)

As access to higher education has expanded, understanding the influence of socioeconomic status (SES) on educational outcomes has become increasingly important. Studies have shown that low-income students are less likely to complete bachelor's degrees. Moreover, completion rates differ drastically by race: in 2006–07, White students earned 72 percent of all bachelor's degrees, while Black, Hispanic, and Asian/Pacific Islander students' degree attainment was significantly lower. Given the United States' changing demographics, these trends are troubling – making it even more essential that survey research is conducted that correctly identifies how SES impacts postsecondary success.

Current research uses various proxies for SES – maternal education and parental occupation being two of the most popular. Although debate continues about whether a single or composite measure best represents SES, it is essential that researchers strive to ensure that their measure of SES is an accurate representation of the socioeconomic backgrounds of their respondents. However, discrepancies between student-reported and actual parental occupation and education level persist. This is particularly troublesome because many surveyors cannot verify student self-reporting, relying on data collected from student interviews to determine SES.

This paper addresses this issue by using the National Longitudinal Survey of Freshmen (NLSF) to evaluate whether internal inconsistencies between student self-reports of parental education and occupation impact estimates of the effect of college persistence and graduation. OLS and logistic regression analyses will be the fundamental analytic tools used to answer these questions. Finally, because minority students were oversampled in the NLSF, I will also be able to examine differences by race. One implication of this research is that previous studies may have over- or under-stated the impact of SES on college success.

The Effects of Computer Assisted Recorded Interviewing (CARI) on Questionnaire Timings on the NLSY97.

Jodie A. Daquilanea, National Opinion Research Center at the University of Chicago (NORC) (daquilanea-jodie@norc.org); Arleigh Dom, National Opinion Research Center at the University of Chicago (NORC) (dom-arleigh@norc.org); Kanru Xia, National Opinion Research Center at the University of Chicago (NORC) (xia-kanru@norc.org)

The National Longitudinal Survey of Youth 1997 (NLSY97) provides a suitable environment in which to test the effects of audio recording on interview administration timings. The NLSY97 is a national long-term labor market survey sponsored by the Bureau of Labor Statistics in which the same respondents, born in 1980 to 1985, are interviewed annually. On the twelfth round of the survey, Computer Assisted Recorded Interviewing (CARI) was implemented for all field interviewers. Selected questions were recorded for all interviews in which the respondent consented to audio recording, and interviewers were provided with feedback on their interviewing performance based on review of the recordings of the interviews they conducted. Interviewers were not told which questions in the interview were recorded. Since there was a great deal of emphasis during Round 12 on the use of audio recordings to monitor interviewer behavior, we conjectured that interviewers would more consistently focus on such aspects of performance as reading the questions as worded and probing thoroughly. We expected that increased attention to these qualitative aspects of performance would result in increased interview administration time.

Initial results indicate that the average duration of Round 12 interviews increased by about 10% from Round 11 although approximately the same number of questions were asked in Round 12 as in Round 11. When we looked at the interview administration time by question, there was about a 10% increase in administration time. An even higher increase in questionnaire administration timing was apparent when comparing the timings of introductory questions; for example, some of the questions showed a 29% increase of administration time. In this paper we will explore the impact that audio recording has upon the interview experience. These findings will add to the body of knowledge about the effects that audio recording of interviews has upon survey metadata.

Minimizing Survey Non-response

The Road to Understanding Nonresponse in the 2008 June Area Survey

Michael Gerling, USDA-NASS (michael_gerling@nass.usda.gov); Terry O'Connor, USDA-NASS (terry_oconnor@nass.usda.gov); HoaiNam Tran, USDA-NASS (hoainam_tran@nass.usda.gov)

The National Agricultural Statistics Service (NASS) surveys farmers and ranchers across the United States and Puerto Rico in order to estimate crops and livestock, assess production practices and identify economic trends. One of the surveys conducted annually is the June Area Survey (JAS). This survey requires NASS' field interviewers (called field enumerators) to visit sampled land areas (segments) designated on aerial photos and record all agricultural activity occurring within those specified land areas. From 2000 to 2007, the national JAS's overall response rate had been gradually deteriorating. In 2000, the response rate was 86.5 percent and had fallen to 81.7 percent in 2007. If this trend continues, the JAS national response rate would fall below the Office of Management and Budget's threshold rate of 80 percent in three to four years. Falling below this rate dictates the need for nonresponse bias analysis and, in general, heightens the concern about the potential negative impact on survey results. There is a clear need to address nonresponse issues on this survey.

This study examines some of the underlying causes of nonresponse occurring in the 2008 June Area Survey.

We Want You and Your Colleague: An Experiment to Assess Survey Design Features that Maximize Response Rates Among Nursing Home Providers.

Melissa Clark, Brown University, Program in Public Health (melissa_clark@brown.edu); Michelle Rogers, Brown University, Program in Public Health (Michelle_Rogers@brown.edu); Andrew Foster, Brown University, Department of Economics (Andrew_Foster@brown.edu); Faye Dvorchak, Brown University, Program in Public Health (Faye_Dvorchak@brown.edu); Frances Saadeh, Brown University, Program in Public Health (Frances_Saadeh@brown.edu); Jessica Weaver, Brown University, Program in Public Health (Jessica_Weaver@brown.edu); Vincent Mor, Brown University, Program in Public Health (Vincent_Mor@brown.edu)

Directors of Nursing (DoN) and Administrators (ADMIN) play important roles in the leadership of long-term care facilities. Perspectives from both individuals are often important in understanding the clinical and administrative issues faced by these institutions. We conducted an experiment to assess ways to maximize the likelihood of obtaining completed questionnaires from both the DoN and ADMIN at a facility. Facilities were stratified based on number of facilities per state, ownership type, bed size, and percent of non-white residents. A total of 224 facilities were randomly selected from these strata and allocated to 16 conditions based on data collection mode (web vs. mail), questionnaire length (short vs. long) and incentive condition. Four incentive conditions were based on amount compensated if the individual completed and an additional amount per individual if the pair completed: (1) \$30 individual/\$5 pair/\$35 total; (2) \$10 individual/\$25 pair/\$35 total; (3) \$30 individual/\$20 pair/\$50 total; (4) \$10 individual/\$40 pair/\$50 total. Each respondent in the facility received an introductory packet, a mailed reminder letter, and an average of 6 telephone follow-up contacts over eight weeks. Overall, 47.4% of eligible respondents completed the questionnaire (45.3% DoN; 49.5% ADMIN). Both respondents completed the questionnaire in 29.3% of facilities. In multivariable analyses, there were no differences in the likelihood of both respondents completing the questionnaire by mode, questionnaire length, or incentive condition. The likelihood of both individuals responding was higher in facilities with fewer minority residents (AOR=1.9, 95% CI=1.0, 3.6) and lower for free-standing, for-profit facilities vs. hospital-based facilities (AOR=0.4, 95% CI=0.2, 0.9). Average numbers of contact attempts were higher for the web mode, long questionnaire, and \$10/\$40/\$50 incentive condition. Higher compensation amounts contingent on completion by both members of the pair was not an effective strategy for increasing the likelihood of both members of the facility participating in the study.

Analysis of Non-Response Bias in a Survey of Newly Naturalized Citizens

Kelly Daley, Abt SRBI (k.daley@srbi.com); Allison Ackermann, Abt SRBI (ackermann@srbi.com)

Abt SRBI designed and fielded the Survey of Recently Naturalized Citizens from October 2009 to February 2010, employing three sequential modes of data collection in order to achieve a targeted 70% response rate. The strategy included intense respondent tracing, a rigorous telephone effort (including bi-lingual interviews in six languages and third party interpreters), an abridged mail survey and finally, in-person recruitment attempts in select cities nationwide. In this paper we present the results of each level of effort for survey response bias and cost per completed interview. We will compare respondents and non-respondents on a wide range of variables that are available in the FY 2008 sampling frame data base, including age group, date of naturalization, gender, country of birth, Census Region, degree of urbanicity of the county of residence. We expect that individuals who required in-person interviews to have similar characteristics to those individuals that we never interview. We will therefore compare

individuals interviewed by telephone or mail with those interviewed in-person in terms of substantive questionnaire variables.

The survey of Recently Naturalized Citizens is a national survey of 5000 adults who became naturalized U.S. citizens during fiscal year 2007-2008. A two-stage sampling design was deployed, drawing from N-400 application for citizenship database provided by the United States Citizenship and Immigration Services. The survey is designed to provide an in-depth profile of the characteristics of naturalized immigrants, and importantly, to uncover motivations behind the decision to naturalize and the timing of that decision. Data from the survey will help the United States Citizenship and Immigration Services to improve projections of the immigrant naturalization applications.

A Comparison of Non-Recruited, Non-Responding and Responding 18-24 Year Old Web-Based Panel Members' Exposure to and Uses of New and Emerging Smokeless Tobacco Products
Trent D Buskirk, Saint Louis University (tbuskirk@slu.edu)

Nonresponse bias in surveys has been well documented in the literature for many substantive questions of interest. Cost, time and other factors though often prevent a more thorough examination of the nature of the nonresponse bias. In this study we report the efforts of a comprehensive examination of both demographic and smokeless tobacco product usage and advertising exposure among 18-24 year olds who were recruited as part of a web-enabled panel selected via a national probability sample using an address-based frame.

While generally small, there is often panel non-response for any given survey administered to a panel. However persons within the 18-24 year old age range are typically a harder to reach segment of the population. To better understand this portion of non-response, this study collected comprehensive demographic information from panelists who did and did not respond to the smokeless tobacco survey (non-responding panelists). A comparison of these demographic variables between the panel survey respondents and non-respondents provides the first part of the reported results.

Because panel surveys generally have higher participation rates (high responses among those panel members on any given survey) but lower recruitment rates, this nonresponse study also conducted a follow-up survey for a random sub-sample of 18-24 year olds who were non-respondents of the original panel recruitment efforts (non-responding, non-recruited panelists). The non-response survey contained a shortened version of the original survey that focused on three key areas education, use of smokeless tobacco products and exposure to tobacco advertisements. This follow-up survey was administered using a comparable mode with a more aggressive recruitment strategy (phone and email). A comparison of these outcomes as well as basic demographic information to those of the responding panelists forms the second part of the reported results.

How Random is an Online Respondent Pool? Analysis of Respondent Self Selection in Multiple Surveys to a Large Online Volunteer Panel

Steven Millman, Booz Allen Hamilton (millman_steven@bah.com); John Zogby, Zogby, International (john@zogby.com); Zeljka Buturovic-Bradaric, Zogby, Int'l (zeljka@zogby.com); Katherine Schwalbe, Zogby Int'l (katy@zogby.com)

In this study (primary data collection in progress), the investigator team seeks to examine the possibility of one form of self-selection bias among panel members in a large online volunteer panel. Specifically, we seek to investigate the criticism that panel members may regularly respond more than others, and that those who respond regularly to surveys hold beliefs that are different – by perhaps being more extreme or

intense - that make them respond to surveys more frequently while simultaneously introducing bias into results. It should be noted that some studies by other investigators suggest that regular responding to internet surveys does not by itself introduce significant bias into the results. This might be related to the finding that the main motivation of regular respondents appear to be that of duty or desire to help rather than express one's views. In our ongoing project, we will be sending a total of three surveys addressing political, economic and election issues to the same randomly selected group of potential respondents from the Zogby Interactive Panel. These surveys will feature very similar questions in addition to gauging one's motivation in responding to survey questions more generally. Different groups of respondents will be identified depending on the number of surveys they completed. We will then make comparisons among these different groups of respondents regarding the substantive views that they have expressed in order to identify the degree to which the individuals' frequency of responding to surveys affects his answers to survey questions. In addition, we will compare demographics for these different groups to identify any stable demographic differences among groups of panel members with different propensities of completing surveys.

Questionnaire Design: Context, Placement & Length

Health Care Reform Perceptions: Context and Reciprocity Effects.

Peter Graven, SHADAC (grave165@umn.edu); Todd Rockwood, University of Minnesota (rockw001@umn.edu)

The current national debate over health care reform routinely focuses on two issues: finance and access. This paper will report on a series of experiments which targeted question order and wording affects in the assessment of these two issues. These experiments have been embedded in the Robert Wood Johnson Foundation Health Care Opinion section of the Consumer Sentiment survey. To date, 1,001 surveys have been completed, though data collection will continue at a rate of 500 per month through the conference date. The order experiment seeks to test if people answer about the impact of reform differently if they are asked in reference to themselves or the country first. Overall, the question shows that while just 17 percent believe reforms will make their own access better, 38 percent believe it will make the country's access better (a 22 point difference). Within these results the data demonstrates a norm of even-handedness, or reciprocity effect, for access in which the responses to the 'self' item remain relatively constant regardless of order, just a 1 point difference between when asked about self 1st, compared to the 'country' item in which there is a shift of 11 points for whether it is asked first or after self. Furthermore, this effect is constant regardless of whether the 'make things better' or 'make things worse' appears first in the question stem. The full paper will consider interaction effects between the experiments, update the analysis with the most recent data and consider separately effects on the "better" "same" and "worse" responses.

Using Item Response Theory and Computer Adaptive Testing to Reduce Survey Length in Opinion and Attitude Surveys

Scott Hershberger, Harris Interactive (shershberger@harrisinteractive.com)

Applications of item response theory (IRT) have increased considerably because it (a) provides more in-depth analysis of items included in measurement instruments, (b) allows instrument developers to better handle complex measurement problems such as linking scores from various instruments, and (c) facilitates the development of more efficient instruments by reducing the number of questions to be included in a scale. This presentation focuses on (c), IRT's ability to guide the construction of shorter surveys of equal or, paradoxically, greater reliability than methods based on more traditional

measurement models (e.g., classical test theory). IRT accomplishes this through computer adaptive testing (CAT). In CATs, a large pool of questions, called an item bank, is used that have been calibrated by IRT. Based on a respondent's response to an initial question, the CAT chooses the item bank's next most informative question to administer to the respondent. After each response, the computer selects and administers from the item bank the next most informative question until a reliable score is obtained. In this way, CATs, in conjunction with IRT, tailor a survey to a particular respondent, this tailoring based on a respondent's standing on whatever behavior, trait or attitude is being measured. This presentation provides details about this methodology, as well as a case study where an opinion and attitude survey was successfully administered using IRT to develop the item bank and CAT to administer the survey.

A Comparison of Results from a Spanish and English Mail Survey: Effects of Instruction Placement on Item Missingness.

Kevin Wang, RTI International (kwang@rti.org); M. Mandy Sha, RTI International (msha@rti.org); Emilia Peytcheva, RTI International (epeytcheva@rti.org)

Few research studies compare results from self-administered bilingual paper questionnaires on how the positioning of skip instructions may affect the respondent's ability to follow skip patterns. Using data from the 2004 and 2005 of the Phase 5 Pregnancy Risk Assessment Monitoring System (PRAMS) mail survey questionnaire, this paper attempts to fill this gap. We examined whether the positioning of skip instructions can produce differences in item nonresponse rates in subsequent items and how results compare between English and Spanish language questionnaires.

About 30 PRAMS vital records registry areas (consisting of 29 states and New York city) had their own mail questionnaire in 2004-2005, and the majority also offered a Spanish language version. Because different questions could be selected by each PRAMS area for certain question topics and the questionnaire layout is in a two-column multi-page format, there were five different placements for the skip instruction for a common item across the surveys. Our goal is to examine how differences in displays of skip instructions were associated with differences in item nonresponse. We found that across all conditions there is higher data missingness in the data collected through the Spanish questionnaires as compared to the English questionnaires for both Non-Hispanics and Hispanics. However, the type of skip instruction placement has more of an effect on item nonresponse in the English version than in the Spanish version. These results will be of interest to designers of bilingual self-administered survey questionnaires in guiding respondents through the intended navigational path with skip patterns.

Question Order Effect on Self-Rated General Health in a Multilingual Survey.

Sunghee Lee, UCLA Center For Health Policy Research (slee9@ucla.edu); David Grant, UCLA Center for Health Policy Research (dgrant@ucla.edu)

In order to minimize order effects, current public health literature recommends that specific health status items follow a general, self-rated health (SRH) question in survey questionnaires. Contrary to this recommendation, a recent study (Lee and Grant, 2009) found a significant and substantial interaction between the order of SRH in a questionnaire and interview language: more than half of the Spanish-language respondents reported fair/poor health when SRH was asked preceding chronic condition questions, whereas only one out of three reported fair/poor health when SRH was asked following chronic conditions (adjusted for age and gender). This effect was not found in English-language interviews. To further explore the interaction of question order and language on SRH, we implemented a split ballot experiment in a health survey where SRH was asked in four different locations: 1) as the first among all health related questions; 2) after a set of chronic condition questions; 3) after a set of mental health questions; and 4) at the end of all health-related questions. We examine the effect of different locations of

SRH by interview language, the relationship between location of SRH and its preceding content, and the potential consequences of the order effect in health disparity research findings.

Putting Context Effects in Context: The Variable Effects of Question Order

Michael Dimock, Pew Research Center for the People & the Press (mdimock@pewresearch.org); Shawn Neidorf, Pew Research Center for the People & the Press (sneidorf@pewresearch.org); Leah Christian, PEW Research Center for the People & the Press (lchristian@pewresearch.org); Jacob Poushter, PEW Research Center for the People & the Press (mdimock@pewresearch.org)

Much of the research on question order effects – the way in which survey estimates are influenced by preceding questions – has been based on experiments embedded within specific, one-time, data collection efforts. This paper assesses whether the magnitude and direction of context effects vary over time by analyzing three sets of questions that have appeared in random order for many years. Our hypothesis is that question order effects themselves have a context: The extent to which answers to a previous question influence answers to a subsequent question depends on the external conditions in which the questions are asked.

Surveys in the latter part of George W. Bush's second term in office found that respondents who were first asked to rate Bush's job performance consistently reported greater dissatisfaction with national conditions than people who were asked about the state of the nation first, and the magnitude of the difference increased as Bush's approval ratings declined. But over Obama's first year, the effect has changed such that people's dissatisfaction with national conditions have a negative effect on ratings of Obama's job performance. We conduct similar evaluations of question order effects over time on the relative ratings of the job performance of Republican and Democratic leaders in Congress and of national and personal economic ratings. In all three cases, we aim to understand whether and how substantial shifts in the balance of public opinion have changed the way in which question order effects occur.

One of the most important implications of this research is to encourage surveyors to continuously evaluate the effect of preceding questions on people's responses. The presence, or absence, of question order effects may not be constant over time and the magnitude and even direction of the effects can vary over time as political, economic or other conditions change.

Survey Interviewing & Data Quality

A Common, Mode-Independent, Approach For Evaluating Interview Quality and Interviewer Performance: Lessons Learned

Howard Speizer, RTI International (hspeizer@rti.org); Susan Kinsey, RTI International (shk@rti.org); Richard Heman-Ackah, RTI International (rackah@rti.org)

RTI has developed a common, mode-independent interview quality monitoring evaluation system. In-person and telephone interviewer behaviors are evaluated using a common set of quality metrics and interviewer performance data are stored in a single shared database. The system supports this evaluation both in real-time (live monitoring) and through audio recordings of the survey process.

The new system (QUEST) replaces a more variable set of interviewer quality monitoring activities implemented at RTI and for the first time combines interviewer evaluation data across multiple projects. In the first part of this paper we describe the challenges related to designing and introducing this new

system. We then provide insights from the common behaviors database and compare some performance metrics across survey modes and projects.

The emphasis in QUEST on recorded interviewer-respondent interactions provides a richer set of data for monitoring quality and assessing individual performance. The audio recordings can serve as direct and concrete feedback for interviewers who can hear in their own voices the behaviors noted by quality monitors. In the second part of this paper we describe the benefits and drawbacks that we have experienced using audio recordings and the challenges that we have faced as we start using the system for evaluating telephone interviewers that have traditionally been monitored in real-time.

Our ultimate goal is to use the quality monitoring data collected over a large number and wide variety of projects to help focus our quality improvement efforts. In the last part of the paper we discuss our efforts to use performance metrics to improve techniques and approaches being used to train interviewers and the usability of computer assisted survey instruments.

A Two-stage Interviewing Experiment in an RDD Survey.

Sherman Edwards, Westat (ShermEdwards@westat.com); Michael Brick, Westat (MikeBrick@westat.com); Royce Park, UCLA Center for Health Policy Research (npark@ucla.edu); David Grant, UCLA Center for Health Policy Research (dgrant@ucla.edu)

As RDD survey response rates continue to decline, it may be useful to question traditional wisdom about the interviewing process. This paper will describe an experiment in which interviewers on the 2009 California Health Interview Survey (CHIS) were assigned either (1) solely screening interviews or (2) solely extended interviews with sampled adults. Preliminary results indicate that this two-stage approach increased the sample yield when compared with a control group of cases assigned to interviewers in the traditional one-stage mode, where the interviewer attempts to complete the screener and extended interview on the same call.

CHIS is a biennial survey of California households, with one adult selected at random from each household screened. One child and one adolescent associated with that adult may also be sampled. The adult interview is approximately 35 minutes long. For the experiment, groups of telephone numbers were randomly assigned to either the control or experimental treatment. A group of experienced interviewers was sorted by historical screener cooperation rates on RDD surveys, and alternately assigned to the treatment or control condition. In the experimental treatment, those with the highest cooperation rates were assigned to conduct screeners, and the remainder to conduct extended interviews. The experiment was conducted with landline numbers only, and only in English. Refusal conversion was attempted at both the screener and extended level for all cases. All telephone numbers with matched addresses were sent a prenotification letter with a \$2 bill. A further experiment will test the effects of sending a second letter with \$5 to the sampled adult after the screener.

Outcome variables will include the screener and adult interview response rates, as well as the level of effort per completed adult interview. The paper will also discuss considerations in implementing two-stage interviewing for the entire CHIS sample and for other RDD surveys.

Data Quality in the Retrospective Reporting of Addresses

Kate E Bachtell, NORC at the University of Chicago (dalton-kate@norc.org); Virginia Tangel, NORC at the University of Chicago (tangel-virginia@norc.org); Ned English, NORC at the University of Chicago (english-ned@norc.org); Michael Latterner, NORC at the University of Chicago (latterner-michael@norc.org); Catherine Haggerty, NORC at the University of Chicago (haggerty-cathy@norc.org)

While tracking the movement of respondents has always been crucial for panel studies, the increasing popularity of geographic analyses has furthered the demand for both accurate and systematic address collection. This paper advances the existing literature on retrospective reporting in surveys by focusing on the collection of respondents' past addresses. It features data from the third wave of Making Connections, a ten-year, neighborhood-based survey funded by the Annie E. Casey Foundation. The wave three questionnaire featured a new series that solicited a detailed history of respondents' movement during the previous three years. Recovering previous addresses presented challenges beyond those typically associated with the systematic recording of physical addresses because recalling information from the past is inevitably more difficult than describing the present (Kennickell and Starr-McCluer 1997).

We designed an experiment to test different methods of maximizing data quality in the retrospective address series for our 2010 data collection in Providence, Rhode Island. First, we redesigned our interviewer training to include more in-depth exercises that demonstrate how to best elicit and record systematic address history information. The data collected in Providence was compared to the data collected in the wave three sites completed prior to 2010 to determine if the advanced training resulted in higher data quality. In addition, households were randomly assigned to three groups: a treatment group whose addresses underwent administrative data cleaning (using Google Maps, Accurint, etc.); a second treatment group that received intensive follow-up calls or in-person visits by field interviewers; and a control group that received no special post data collection treatment. We investigate the efficacy of each method in producing addresses that can be successfully translated into geographic coordinates and thus used for spatial analyses. Our findings may be used to inform improved methodologies for collecting retrospective data in both panel and cross-sectional surveys.

An Examination of Within-Interviewer Continuity and Change in the use of Standardized Interviewing Following Repeated Corrective Feedback

Kristine Fahrney, RTI International (fahrney@rti.org); Danielle Cohen, RTI International (dcohen@rti.org); Ann Burke, RTI International (aburke@rti.org)

Although standardized interviewing practices are widely advocated in order to reduce interviewer-related measurement error, audio recordings of interviewers' work reveal interviewers' sometimes persistent tendency to paraphrase or reword questions. This presentation includes interviewer performance data obtained through CARI (computer recorded audio interviewing) monitoring on two separate face-to-face surveys. The first is a mid-sized study that employs approximately 85 interviewers who administer an approximately 60-minute computer assisted interview. Most interviewers were monitored at least 3 times during the 3 month period of survey production. Our analysis of this data will focus specifically on quantifying the proportion of interviewers who show improvement in reading questions verbatim after receiving corrective feedback and identifying correlates of responsive interviewers. The second survey employs a small number of interviewers (~13) who administered a complex 90-minute computer assisted interview. Interviewers received feedback 3 times over the course of 6 months. The small size of this team allowed us to capture rich descriptive data on whether or not there was improvement in their use of standardized interviewing techniques following corrective feedback. A brief anonymous survey administered to these interviewers sheds light on competing priorities (e.g. need to follow protocols vs. need to be mindful of the respondent's time or interest level) and is suggestive of ways that researchers can frame corrective feedback to be more persuasive to field staff.

Utilizing Distance Learning to Deliver Standardized, Multi-Media Field Interviewer Training and Reduce Costs on Large-Scale National Field Surveys

Allison McKamey, RTI International (mckamey@rti.org); Scott Payne, RTI International (spayne@rti.org); Elizabeth Shuey, RTI International (eshuey@rti.org)

Comprehensive and standardized field interviewer (FI) training programs are an essential component to successfully conducting large-scale national field surveys. However, with geographically dispersed interviewers and the steady rise in costs associated with travel, traditional in-person training sessions can be financially challenging.

Distance learning offers an exciting and innovative alternative to costly in-person trainings. This presentation will examine the utilization of a distance learning application developed for training FIs on a large-scale national field survey currently conducted by RTI International. This training application, known as iLearning (or Independent Learning), enables FIs to view high-quality training presentations at their own pace on project-provided laptops. The iLearning application incorporates multi-media training tools, such as audio, video, complex graphics, and interactive assessment questions and exercises to deliver standardized, high quality training programs on a wide variety of subjects.

Distance learning applications, such as iLearning, offer the flexibility to train newly-hired FIs on interviewing and research fundamentals, provide refresher training to veteran FIs on project protocols throughout the year, and present ad hoc training for special studies or field tests without incurring the expenses associated with conducting in-person training. Examples showcasing features of the iLearning application and how they are used to present training topics via distance learning will also be discussed.

Finally, this presentation will review sources of potential cost savings associated with utilizing distance learning on large-scale surveys. Implementation considerations and future enhancements will be provided to demonstrate how distance learning applications, such as iLearning, have the capacity and flexibility to present FIs with an effective and enjoyable learning experience that complements traditional in-person training.

AAPOR Concurrent Session B

Saturday, May 15, 2010

10:00 a.m. - 11:30 a.m.

Cell Phone Surveys: Population Estimates, Dual Frames & Weighting Issues

The Cell Phone Universe: Methodological Considerations for Creating Cellular and Landline Telephone Estimates from Nielsen Television Surveys.

Christine Pierce, The Nielsen Company, Media (christine.pierce@nielsen.com); Jonathan D. Stringfield, The Nielsen Company, Media (jonathan.stringfield@nielsen.com); Misty Saline, The Nielsen Company, Media (misty.saline@nielsen.com); Jennifer Hsia, The Nielsen Company, Media (jennifer.hsia@nielsen.com)

The proliferation of cellular telephones in recent years has presented a challenge for public opinion researchers, particularly in regards to coverage error in random digit dial (RDD) data collection and telephone frame sampling techniques that are increasingly unable to reach households that use cell phones as their exclusive telephone line. This challenge has established a need for current estimates of cell phone-only households to evaluate potential biases resulting from undercoverage, in addition to sample weighting and recruitment. To date, estimates of cell phone-only households have largely been available only periodically and for capacious geographies. Given that an increasing number of households are becoming cell phone-only at a heightened rate, and large-area estimates are of limited utility for many applications, there is a need for household telephone status estimates available in a timely manner for smaller geographic areas based upon contemporary data. To address this need within The Nielsen Company, the Media-demography group is conducting evaluation research on the possibility of creating quarterly estimates of households by cellular/landline telephone status. These estimates would be created by adapting pre-existing methodologies used to create universe estimates for the quickly changing household media environment to the evolving cell phone adaptation environment using inputs from the Nielsen television panels and surveys, data sources that are ideally situated to estimate these differential and difficult to measure household types.

This paper provides a brief overview of media-related universe estimation, discusses limitations and advantages in applying this methodology to estimates of household telephone status, and provides preliminary data on these telephone household types using the Nielsen television data. The results of this research will provide additional insight to the challenges of estimating households of varying telephone statuses and provide alternative and current information on the status of cell phone-only adaptation in the U.S for national and sub-national geographies.

Can Post-Stratification Adjustments Do Enough to Reduce Bias in Telephone Surveys that Do Not Sample Cell Phones? It Depends.

Kathleen Thiede Call, University of Minnesota (callx001@umn.edu); Michael Davern, NORC (Davern-Michael@norc.org); Michel Boudreaux, University of Minnesota (boudr019@umn.edu); Pamela Jo Johnson, University of Minnesota (johns245@umn.edu); Justine Nelson, University of Minnesota (nels2177@umn.edu)

Good sample coverage is a persistent challenge for traditional landline Random Digit Dial (RDD) surveys given the rapid increase in cell phone only households (CPOH). Yet, common challenges to CPOH surveys are their expense and complexity of merging the CPOH and RDD cases. We examine how biased health surveys are when they omit CPHOs, explore whether post-stratification can reasonably reduce this bias, and consider how well these adjustments work for key subpopulations (young adults and minorities). We use 2008 National Health Interview Survey (NHIS) data which includes both types of households and survey data on CPOH status. First, we replicate earlier findings showing that people living in CPOH are different from those who do not live in CPOH with respect to several important health surveillance domains (e.g., health insurance coverage, access to care, smoking and drinking). Then we compare standard NHIS estimates to a set of “re-weighted” estimates that exclude people living in CPHO. The “re-weighted” cases are fitted through a series of post-stratification adjustments to NHIS control totals. We select the weight that performs best based on the mean squared error (MSE) for each outcome. In addition to the usual regional and demographic post-stratification adjustments, we demonstrate the importance of including an adjustment for home ownership status. Post-stratification reduces bias in all health related estimates for the total and non-elderly population. However, these adjustments work less well for Hispanics and Blacks and even worse for young adults (18-30). Generally, reduction in bias is greatest for estimates of uninsurance and having no usual source of care and worse for estimates of drinking, smoking, and forgone or delayed care due to costs. We conclude that post-stratification adjustments may not do enough to reduce bias in health related estimates, particularly for those interested in measuring and monitoring racial, ethnic and age disparities.

First-Stage Weights for Overlapping Dual-Frame Telephone Surveys.

**Jonathan Best, Princeton Survey Research Associates International (jonathan.best@psra.com);
Larry Hugick, Princeton Survey Research Associates International (larry.hugick@psra.com)**

This paper presents recommended first-stage weighting adjustments for surveys administered to overlapping land line and cellular RDD sample frames. The paper will investigate the weighting adjustments necessary to account for unequal probabilities of selection based on the number of adults in each sampled household and the number of telephone lines (either land line or cellular) available to each sampled household/respondent. First-stage weighting adjustments will be derived for both household-level and person-level sample estimates. After the weighting adjustments are derived, we will apply them to recent national dual-frame surveys and examine the effects they have on the demographic makeup of the samples.

Dual-Frame Weighting of RDD & Cell Phone Interviews at the Local Level

Michael P. Battaglia, Abt Associates Inc. (mike_battaglia@abtassoc.com); Donna Eisenhower, NYC DOHMH Epidemiology Services (deisenho@health.nyc.gov); Stephen Immerwahr, NYC DOHMH Epidemiology Services (simmerwa@health.nyc.gov); Kevin Konty, NYC DOHMH Epidemiology Services (kkonty@health.nyc.gov)

Much of the research presented on dual frame surveys has involved national surveys that make use of the National Health Interview Survey estimates of the size of telephone usage groups. The 2008 New York City Community Health Survey consisted of 7,554 interviews with adults in households contacted using a landline RDD sample and a pilot study of 590 cell phone only adults from a cell phone sample frame. How does one go about weighting a local-level, non-overlapping dual frame survey?

Control totals for weighting by telephone usage (cell only, landline only, and dual service adults) were calculated in two ways. Our model-based method fit a household-level multinomial logistic regression model of telephone use to data from the 2007 NHIS, where use group was a categorical dependent

variables and socio-demographic variables were predictors. This model was then applied to New York City data from the 2005-2007 ACS to estimate telephone usage control totals for adults in each of the city's five boroughs. Our second approach used direct estimates of telephone usage groups for each borough from the 2008 NYC Housing and Vacancy Survey, an in-person survey of 18,000 households conducted by the Census Bureau every three years. HVS data suggested a larger cell phone only population (18%) than did the NHIS model (15%).

We then compared weighted estimates for seven different health risk factors and conditions. Differences in overall and subgroup prevalence were small and rarely rose to a level of statistical significance. The results suggest that weighting to local independent estimates is more accurate than the model-based approach, but that even subgroup differences in resulting prevalence estimates are generally small. Implications for state and local dual frame surveys are discussed.

Assessment of Alternative Methods to Compensate for Noncoverage of Nonlandline Telephone Households in the NIS

Meena Khare, NCHS/CDC (mkhare@cdc.gov); Kennon R. Copeland, NORC at the University of Chicago (Copeland-Kennon@norc.org); Nadarajasundaram Ganesh, NORC at the University of Chicago (Ganesh-Nada@norc.org); Abera Wouhib, NCHS/CDC (Awouhib@cdc.gov); James A. Singleton, Centers for Disease Control and Prevention (CDC), NCIRD (JSingleton@cdc.gov)

Random-digit-dialing (RDD) surveys are subject to bias due to noncoverage of households with no access to landline telephones ("nonlandline households") – both households with access only to wireless telephones ("wireless-only households") and households with no telephone service ("nontelephone households"). The proportion of wireless-only households has been increasing over time and is currently estimated at over 20 percent, while the proportion of nontelephone households has been relatively stable at roughly two percent. A common method to adjust for noncoverage of nonlandline households for in RDD surveys has been Keeter's method using information on telephone service interruptions; however this method was designed to compensate for the noncoverage of nontelephone households, not for the increasing noncoverage of wireless-only households.

The National Immunization Survey (NIS) – a nationwide, list-assisted RDD survey sponsored by the Centers for Disease Control and Prevention and fielded by NORC – monitors the vaccination rates of children between the ages of 19 and 35 months and adolescents between the ages of 13 and 17 years. This research examines the performance of alternative methods to compensate for noncoverage of nonlandline households in the NIS, utilizing additional variables collected in the NIS during 2008-2009: household telephone status (i.e., access to wireless telephones) and household tenure (owner/renter). We use three methods of weight adjustments for noncoverage: M1) the usual Keeter's method; M2) a hybrid of the Keeter's method wherein landline-only households are included in conjunction with households with telephone service interruptions; M3) a raking ratio adjustment using socio-demographic population controls correlated with vaccination status and telephone status. Selected NIS estimates are compared to evaluate performance of the three weighting methods and socio-demographic distributions by telephone status are compared to those from the National Health Interview Survey.

Combining Survey & Administrative Data: Issues of Quality & Access

Administrative and Survey Data: An Analysis of Reasons for Differences in Reported Employment Figures.

Polly Phipps, Bureau of Labor Statistics (phipps.polly@bls.gov); Daniell Toth, Bureau of Labor Statistics (toth.daniell@bls.gov)

Two Bureau of Labor Statistics programs - the Quarterly Census of Employment and Wages (QCEW) and the Current Employment Statistics (CES) survey - collect monthly employment figures from an establishment for the same reference period. QCEW data are collected through state unemployment insurance tax filings submitted by establishments, while CES is a monthly sample survey of establishments. The employment values should be identical, but differences between the two sources are observed at the micro level. In addition, seasonal patterns are observed at the aggregate level. In this paper, we describe the employment differences between the two programs, including seasonal patterns. Then we analyze responses to a data quality survey in which 3,000 establishments with large employment differences between the administrative and survey data were asked about reasons for the differences. The data quality questionnaire included separate sections on QCEW and CES reporting, as well as general questions on establishment characteristics and recordkeeping. We focus on respondents' understanding of program definitions and instructions, records availability and use, establishment data sources and recordkeeping practices, as well as specific explanations provided by respondents on reasons for the differences. In the analysis, we account for the complexity of the establishment, including size and multiple locations; and data collection complexity and issues, including multiple respondents, data collection modes, and respondent timeliness. Initial analyses indicate that reporting error occurs in both the administrative and survey data, although the two programs have different patterns and sources of error.

Administrative Data Quality Assessments in the U.S. Federal Statistical System: Current Practice and Problems.

Jenna A Fulton, Joint Program in Survey Methodology, UMD (jfulton@survey.umd.edu); Frauke Kreuter, Joint Program in Survey Methodology, UMD (fkreuter@survey.umd.edu); William Iwig, NASS- USDA (bill_iwig@nass.usda.gov)

The statistical use of administrative records in the Federal statistical system continues to grow. The interest in this type of data, either in lieu of or in conjunction with traditionally collected survey data, is largely motivated by the reduction in data collection costs and respondent burden that administrative records offer. Administrative data are often considered to be of higher quality than traditional survey data, making them an attractive option for assessing bias in and benchmarking survey estimates. Currently, there is little documentation on the methods used within agencies to evaluate and promote the quality of linked and stand-alone administrative records used for statistical purposes. This paper will report on qualitative research conducted with the principal federal statistical agencies that explores current practices in evaluating the quality of administrative records. It will describe how agencies determine adequate quality of record data, the type and frequency of quality assessments they conduct, and any problems and challenges they experience in assessing the quality of their administrative record data.

You Can Match My Data! Biasing Effect in the Use of Linked Administrative and Survey Data.

Stefan Bender, Institute for Employment Research (IAB) (stefan.bender@iab.de); Martina Huber, Institute for Employment Research (IAB) (martina.huber@iab.de); Alexandra Schmucker, Institute for Employment Research (IAB) (alexandra.schmucker@iab.de)

Surveys often cope with special problems: gaps in retrospection appear or respondents could not provide details. Sometimes these problems can be solved by using additional information from process-generated data. These administrative data offer valid and exact information, but include probable less valid variables with a rather high share of missing values, too. By linking the data, the data quality can be improved by

creating a dataset that balances the disadvantages of the administrative and survey data using the advantages of these types of data.

The survey data contain information about 6400 employees from 150 establishments and can be linked via the social security record to administrative data, called the 'Integrated Employment Biographies of the IAB (IEBS)'.

Because the sampling of the survey is done on the basis of the IEBS, the following topics will be analyzed:

1. Unit Nonresponse: because information from the administrative data for all non-participants are available.
2. Selectivity of the match: Because of privacy concerns, the respondents have to agree to the match between their survey data and their administrative data.
3. Deviations: After the comparison of sex and age, which is available in both data sets, we identify deviations and try to correct these variables.
4. Differences in variables: Some variables are available in both data sets (for example: qualification and education). These variables are classified as weak in the administrative data. We will compare the information.
5. Retrospective information: Both data sets contain employment histories. We take a closer look on an overlapping period of one year. Missing episodes in the administrative data can be explained by reported information in the survey data. Vice versa recall errors or missing data in the survey data can be corrected by linking administrative data. Furthermore we show some hypothetical examples of deviations and identify determinants which have an influence.

Using Linked Survey and Administrative Data Files to Improve Surveys and to Evaluate Decentralized Public Programs: The Case of Medicaid and SCHIP Enrollment.

Michael Davern, University of Minnesota (daver004@umn.edu)

Linked survey and administrative data have often been used to assess the quality of survey data self-reported public program enrollment information and income amounts. Another potentially strong contribution these linked data files could make is to help evaluate how well state and local run public programs are serving their clients.

Using a data file with the Current Population Survey's (CPS) cases linked to the Medicaid Statistical Information System (MSIS) we examine whether we detect state differences in reporting patterns of public program enrollment (Calendar years 2000-2003). We generate two logistic regression models predicting whether the CPS case was appropriately coded as having Medicaid in the CPS having Medicaid given that the MSIS shows enrollment. The second model examines whether CPS respondents linked to MSIS showing enrollment in Medicaid answer that they are uninsured in the CPS interview. Both models control for important variables including utilization of services, and state of residence. We feel the state variation in the first model estimating whether a case was correctly classified as having Medicaid could be caused by many things such as stigma and confusion over the exact plan name someone is enrolled in (e.g., SCHIP versus Medicaid). However, we feel an error to report any coverage at all is more serious as it shows people do not know they are enrolled in coverage that the State is providing (and therefore perhaps act more like they are uninsured and have health outcomes more like the uninsured). We demonstrate wide variability among the states in the probability of being coded in the CPS as being uninsured given that the person shows Medicaid enrollment. In this case "survey response error" can also be seen as a programmatic evaluation tool.

Quality and Quantity - Using Administrative Data for Scientific Purposes in Labor Market Research

Patrycja Scioch, Institute for Employment Research - IAB (patrycja.scioch@iab.de); Stefan Bender, Institute for Employment Research - IAB (stefan.bender@iab.de)

Administrative data gain more and more attention in labor market research, in combination with survey data or as sole base data. In contrast to the main advantages like large sample size, long time periods and the extreme wealth of information they suffer of shortcomings like inconsistencies and missing values. Unfortunately the research in quality and the improvement of the data is scarce.

We focus on the education variable in a widely used German administrative data set for labor market research, the Integrated Employment Biographies (IEB), which is very important for analyzes such as wage inequalities and employment chances. The Quality of this variable became extremely worse over the last 10 years, with 35% individuals without any information about the educational attainment. Some studies (Fitzenberger et al. 2006; Drews 2006) tried to improve the quality by developing rules to fill the gaps in dependence on existing information without using additional sources. We use as an additional data source: German patent data for 2002 are linked to verify the data at hand. Approximately 90% of inventors are in the social-security system. Therefore the high-educated are identifiable via their name titles (professor, doctor or diploma).

The education information is compared to the original information in the data and with the outcomes of the imputation rules. An explanatory model quantifies the differences between the inventors and the different education definitions.

Improving Response Rate

The Effects of Tailored Design Method on Response Rates in a Chronically Ill Population: Results of a Controlled Experiment.

Ellen Denzen, National Marrow Donor Program (edenzen@nmdp.org); Todd Rockwood, University of Minnesota (rockw001@umn.edu); Nancy Omondi, National Marrow Donor Program (nomondi@nmdp.org); David Radosevich, University of Minnesota (davidmr@umn.edu); Elizabeth Murphy, National Marrow Donor Program (emurphy@nmdp.org)

Declining response rates to self-administered surveys has led to the development of Tailored Design Method (TDM), a method of survey implementation utilizing design elements such as prepaid incentives, multiple contacts and mixed-modes. Response rates are commonly used to measure survey effectiveness, a controversial measure given that response rates in medical journals average only 62%. This study assessed the effects of TDM on response behavior and nonresponse bias in a survey of patients diagnosed with life-threatening disorders treatable by blood and marrow transplant. The National Marrow Donor Program's Office of Patient Advocacy conducted the Patient Satisfaction Survey (PSS), a cross sectional survey (N=800), in which subjects were randomly assigned to either the control or experimental group. Subjects in the control group received one survey by mail. Subjects in the experimental group received a pre-notice letter followed by mail survey and \$2 prepaid incentive. Those that did not respond to the mail survey received a follow-up mail survey after two weeks. Finally, nonrespondents were contacted by telephone for a follow-up interview. Survey respondents received thank-you post cards. Response rates differed significantly between the control (12.8%) and experimental (50.5%) groups ($p < 0.05$). Subjects in the experimental group (OR 0.16), aged more than 50 years (OR 0.52) and having "other diagnosis" (e.g. hematological diseases, immunological deficiencies and genetic disorders; OR 0.56) were more likely to respond. Using TDM in chronically ill populations improves response rates and enhances demographic representativeness thereby reducing nonresponse bias of survey results.

Mail Surveys: The Power of the Return Address.

Larry Luskin, ICF Macro (lawrence.a.luskin@macrointernational.com); Sherri Settle, ICF Macro (sherri.a.settle@macrointernational.com)

The use of communication pieces (pre-notification letters, reminder letters, etc.) has proven to be an effective way to increase response rates in mail surveys. However, these communication pieces are not effective if an individual tosses them as junk mail before opening them. The return address on a piece of mail is what an individual looks at when determining when to open the mail or throw it out unopened. The Internal Revenue Service (IRS) and ICF Macro recently redesigned the survey mailing materials to clarify the federal sponsorship and contractor role. Previously, the return address on the envelopes displayed the data collection center address without reference to the federal agency sponsoring the survey. The envelopes now include both IRS and ICF Macro in the return address. We will synthesize the results of seven recent mail based studies to determine whether this redesign has a positive, negative, or no effect on the survey response rate.

Methods for Elicitation of Prior Information on Survey Data Quality and Survey Cost Structures
John L. Eltinge, Bureau of Labor Statistics (Eltinge.John@bls.gov)

Practical work with survey design generally involves a balance among many factors that influence both data quality and cost structures. In recent years, increasing constraints on survey budgets have highlighted the importance of obtaining solid information on the quality-cost balance. Ideally, one would obtain that information with a thorough set of empirical evaluations, including, e.g., cognitive laboratory tests, pilot studies and larger-scale field tests. However, budgetary and time pressures often constrain the amount of initial empirical evaluation that can feasibly be obtained. Consequently, one often uses initial qualitative assessments by field personnel or other experts to identify the modest number of specific options that will subsequently receive more rigorous empirical evaluation.

Many other fields (e.g., medicine, accounting and reliability engineering) historically made similar use of qualitative assessments in the preliminary development of more rigorous empirical work. However, some of those fields have found it useful to complement these initial qualitative approaches with more quantitative methods described as “Bayesian elicitation of prior information.” This paper explores the possible use of similar Bayesian elicitation methods in the preliminary stages of survey design. The development emphasizes: (a) the conceptual background for Bayesian elicitation methods; (b) connections between these methods and the psychology literature; (c) related previous work in the survey methodology literature; (d) specific components of survey data-quality and cost-structure evaluations that may be compatible with these elicitation methods; and (e) links between (d) and the features of successful previous applications in other fields.

Key words: Bayesian prior distribution, cognitive laboratory study, field test, fixed and variable cost components, nonresponse, nonsampling error, pilot study, total survey error

The SNAAP Experiments: An Attempt to Improve Response and Measure Error

John Kennedy, Indiana University Center for Survey Research (kennedyj@indiana.edu); Steven Tepper, Vanderbilt University (steven.j.tepper@Vanderbilt.Edu); Sally Gaskill, Indiana University (gaskill@indiana.edu); George Kuh, Indiana University (kuh@indiana.edu)

The Strategic National Arts Alumni Project (SNAAP) surveys alumni from arts high schools, arts colleges, and arts departments within colleges and universities. The focus of SNAAP's Internet survey is to examine the career outcomes of their graduates (snaap.iub.edu). The target populations are alumni who

graduated 5, 10, 15, and 20 years prior. SNAAP was designed to have three pilot administrations. This presentation reports on experiments conducted as part of the second pilot study which surveyed alumni from 54 arts institutions.

In the initial pilot administration, the survey was hampered both by a lack of accurate email addresses and low response rates. For the second pilot administration, two experiments were conducted to attempt to improve response and to help determine the impact of the low response rate.

The experiments used alumni from five participating institutions who graduated one year before the targeted cohorts. One experiment provided a contingent incentive of a \$15 gift card emailed immediately upon completion of the questionnaire. A second experiment used a lottery of five \$100 prizes. These incentives were provided to improve response rates and reduce abandonment.

A third experiment used a short version of the questionnaire with multi-mode follow-up attempts and a \$20 contingent incentive. Checks were mailed to those who completed in any mode. The alumni were sent three email recruitment messages, followed by a paper questionnaire, followed by an attempt to complete a telephone interview. The goal of this experiment was to determine if the alumni who participated in the more intense follow-up contacts differed from those who responded using the normal survey procedures.

Neither incentive improved response rates – the rates were virtually the same as the main survey. The response rate improved significantly for the multi-mode experiment.

Group and Cultural Influences on Household Survey Response Rate

Peter K. Kwok, NORC (kwok-peter@norc.org); William Hart, NORC (hart-william@norc.org); Hee-Choon Shin, NORC (shinh@uchicago.edu)

The current research explores a model that addresses member-level and household-level response rates in the REACH U.S. (Racial and Ethnic Approaches to Community Health Across the U.S.) survey. In particular, we tested whether the number of members selected within a household reduces overall member-level response rates, but increases household-level response rates. Results supported this hypothesized pattern, and we used additional exploratory analyses to identify factor(s) responsible for the finding that more selected members in a household reduces member-level response rates. Results from these exploratory analyses were mixed, but generally supported the idea that member-level rates are more likely to be reduced when selected members of a household define a single completed interview by another selected member as their own completed interview (e.g., “My wife did it, which is as good as me doing it”). Our discussion focuses on future areas of research suggested by our novel findings and ways to enhance member-level response rates in surveys where multiple members of a household are selected.

Multimode Data Collection: Question Wording & Mode Effects

Measuring Mode Preference: Question Wording and Question Order Experiments

Jolene Smyth, University of Nebraska-Lincoln (jsmyth2@unlnotes.unl.edu); Kristen Olson, University of Nebraska-Lincoln (kolson5@unl.edu)

Mode preference has been a much discussed, but little researched, concept in survey methodology. Conventional wisdom suggests that providing respondents with the mode they prefer will increase data

quality. Although research has found that people's stated mode preference is influenced by the mode in which they are asked, little attention has been paid to the design and wording of the questions themselves. In this paper we report findings from experimental research on the measurement of the concept of mode preference. Specifically, we address four dimensions of mode preference (overall, convenience, enjoyable, and comfort), the order in which these mode preference questions are asked, the effect of including or excluding a "no preference" option, and the mode of administration. We also examine the stability of respondents' stated mode preference over two survey administrations occurring roughly one year apart. The data for this study come from two surveys. In 2008, a listed sample of Nebraska residents age 19 and above was surveyed by telephone in the Nebraska Annual Social Indicators Survey (NASIS). Respondents to this survey were then experimentally followed up by web or mail between July and September, 2009 (N = 556; AAPOR RR1 45%) in the Quality of Life in a Changing Nebraska Survey (QLCN). In addition, respondents to the QLCN were randomly assigned to one of two experimental questionnaires. In Form A, the mode preference questions were identical to those asked in the NASIS. Form B contained experimentally modified versions of the mode preference questions. If stated mode preference is not affected by question wording and order and remains stable over time, then this lends empirical evidence to support the conventional wisdom upon which many are currently making survey design decisions. However, if this is not the case, then considerably more work needs to go into understanding this concept.

Mode Effects on Symptoms Characteristics: The First Follow-up Survey of Chronic Fatigue Syndrome and Chronic Unwellness in Georgia.

Julie A. Pacer, Abt SRBI (pacer@srbi.com); Rebecca Devlin, Abt SRBI (devlin@srbi.com)

Interview mode may affect respondents' answers to sensitive questions and can result in social desirability bias (Tourangeau et al., 2000). Since 1988, Abt Associates has collaborated with the Centers for Disease Control and Prevention (CDC) Chronic Fatigue Syndrome (CFS) Research Program to identify risk factors and biologic markers of CFS to improve identification, evaluation, diagnosis and management of persons with CFS. From 2007-2009, Abt Associates and Abt SRBI conducted the First Follow-up Survey of CFS. This study followed a cohort from the 2004-2005 baseline study which used a random-digit-dialing (RDD) sample to identify and enroll subjects from metropolitan, urban, and rural Georgia populations.

The First Follow-up study consisted of two components: a detailed telephone interview and a one-day clinical evaluation. The telephone interview obtained information on symptoms, in addition to medical history and life experiences. In the clinical evaluation, subjects provided biological specimens, underwent a physical and mental health evaluation, and completed questionnaires concerning symptoms, medical history, health services utilization, early and adult life experiences, and stress, coping and personality traits. The First Follow-up study of CFS presents a unique opportunity to examine symptoms characteristics as reported in a telephone interview, a computer-assisted self-interview, and a physical or mental health evaluation conducted by a physician or psychiatric interviewer. For a subset of respondents who participated in a follow-up study at Emory University, data from a web survey concerning symptoms characteristics will also be examined. The proposed research explores the impact of interview mode on respondents' answers to questions concerning CFS symptoms characteristics, some of which may be considered sensitive. Other non-sensitive symptom characteristics will also be examined to understand the role of mode, if any, on non-sensitive topics.

Mode Effect or Question Wording? Developing Robust Questions for Mixed Mode Surveys

Edith Desiree de Leeuw, Utrecht University (e.d.deleeuw@uu.nl); Annette Scherpenzeel, Centerdata, University of Tilburg (a.c.scherpenzeel@uvt.nl)

We are now at a point in time, where telephone surveys have to face many challenges and where Internet surveys are not yet fully fit to replace them. Mixed-mode surveys are offered as a promising solution. One of the most pressing challenges facing designers of mixed-mode surveys is combining telephone results with those from web surveys. There is considerable evidence that when different modes are used different results are obtained.

Data from different modes may be different because the modes themselves lead to different response processes, or because different questions are employed in different modes (either by tradition or due to mode specific optimization of the questionnaire). For instance, because of the memory problems associated with verbal communication, telephone designers often use unfolding, or polar (end) labeling of response categories.

The goal of this research is to identify means of being able to combine telephone and web survey results effectively, by choosing formats for asking questions that are less affected by mode, than are others. In a series of experiments we test multiple formats for asking questions (e.g. unfolding opinion questions in multiple steps vs. asking a complete question in one-step) within each of the two survey modes. These experiments go beyond many previous tests because they involved not only comparisons across modes of particular formats, but within modes of two different formats.

Members of a high quality, probability-based Internet panel (LISS-panel, Centerdata) were randomly assigned to one of two modes: a computer assisted telephone interview or a web survey. Within each mode the same series of split ballot experiments on question format were conducted. This design enables us to disentangle question format effects from 'pure' mode effects. In addition, the rich background data of the LISS-panel members makes it possible to correct for potential nonresponse differences between the modes.

Estimating Nonresponse- and Mode Effects in a Mixed-Mode Survey

Peter Lugtig, department of Methods and Statistics - Utrecht University (p.lugtig@uu.nl); Gerty Lensvelt-Mulders, University of Humanistics (glm@uvh.nl)

Mode effects are difficult to separate from nonresponse effects in mixed-mode surveys. Mode effects include acquiescence and social desirability effects due to the presence of an interviewer, recency effects in audio surveys and primacy effects in visual surveys (for example see de Leeuw 2005; Dillman et al. 2008; Kreuter, Presser and Tourangeau 2008). However, in everyday mixed-mode surveys we often cannot distinguish mode-effects from effects due to differences in nonresponse and coverage rates that are caused by the use of multiple survey modes (Biemer 2001).

This paper provides a framework for separating mode-effects from nonresponse-effects by matching similar respondents from different survey modes using propensity score matching and expands on an earlier article. (Lensvelt-Mulders, Lugtig and Hubregste 2009). We use data from a multi-mode survey conducted in the Netherlands in 2007. About 14.000 people in our sampling frame were randomly assigned to a survey mode: either Internet or telephone. The internet-respondents were invited by mail. To assess the technique of propensity score matching, respondents from the Internet-survey are matched to participants in an access-panel who completed an identical survey. After matching, we find that differences in the composition of the Internet sample and access-panel can account for differences in our dependent variables; a set of items that ask respondents about perceived environmental hindrance. In the second part of the paper, we assess the occurrence of mode effects in our survey experiment and match the respondents from the telephone- to the Internet-survey. After matching, we can account for differences due to nonresponse effects. The differences we found before matching however persist after

matching. This indicates the occurrence of mode-effects in our telephone and Internet survey for the environmental hindrance questions. We conclude with a discussion of future mixed-mode survey designs, which could give more insight in the situations in which mode-effects occur.

Using a Mixed Method Approach to Explain the Cross-Cultural Differences in Extreme Responding
Meike Morren, Tilburg University (m.morren@uvt.nl); John Gelissen, Tilburg University (J.P.T.M.Gelissen@uvt.nl)

This paper presents an approach which combines quantitative and qualitative methods to analyze why the four largest minorities in the Netherlands – Turks, Moroccans, Antilleans and Surinamese – respond differently to items such as ‘A man and woman are allowed to live together without being married’. Respondents from culturally diverse groups may have different reasons to answer ‘totally disagree’ on a five-point Likert scale – apart from the traditional family values measured by the item. For instance, Moroccans may not even consider other response categories because of their Islamic background whereas Turks may respond extremely because they defend their honor. To investigate the cultural diversity in using response scales and how it confounds group differences in attitudes, we propose a mixed method approach that combines statistical modeling and cognitive interviewing.

This approach is as follows: We first use a Latent Class Factor model to examine the attitudes across cultural groups while controlling for group differences in extreme response style (ERS). This is done based on large-scale survey data. Then, in cognitive interviews members of the four cultural groups are asked to answer to the same items used in the quantitative model.

The quantitative model is integrated with the cognitive interviews by relating the model estimates to the answers given by the interviewees in the cognitive interviews. This enables us to assess each interviewee’s attitude and response style according to the model. By comparing the arguments given by the interviewees who differ in ERS we gain insight in the response process underlying ERS. Additionally, we posed the same items in different formats (seven-point scale, and vignettes) to investigate whether and how the format of the Likert scale contributes to the cross-cultural differences in extreme responding.

Multimode Surveys: Reliability & Data Quality

Using Computer Assisted Telephone Interviewing for Reinterview Programs and Its Effect on Data Quality
Chrishelle Lawrence, U.S. Census Bureau (chrishelle.lawrence@census.gov)

Centralized Computer Assisted Telephone Interviewing (CATI) is often thought of as a better choice for the current Census Bureau reinterview programs for both cost and performance reasons. Currently, supervisory field staff conducts both telephone and personal visit reinterviews using Computer Assisted Personal Interviewing (CAPI). The focus of this study was to determine the impact on nonresponse bias and respondent error when moving to a telephone only reinterview format. Statistical analysis was completed using reinterview data from the 2008 National Health Interview Survey. Respondents were grouped by reinterview method (personal visit or telephone) and then compared by age, gender, race, and education level. There were significant differences in age, race, and education level between personal visit and telephone reinterview respondents. Minorities, people under age 65, and those without a high school diploma were more likely to be personal visit reinterview respondents. Telephone reinterview respondents were more likely to be White, over age 65, or have a high school diploma. Another issue found among CAPI reinterview cases was the absence of telephone numbers.

Respondents with telephone numbers listed were compared to those without telephone numbers by age, gender, race, and education level. Minorities and people under age 65 were more likely to not have a telephone number listed. All CAPI reinterview respondents were compared to CATI reinterview respondents from the Current Population Survey. There were significantly greater proportions of men and people over age 65 in the group of CATI respondents. Original noninterviews are also verified in reinterview, and are often verified through personal visits. Analysis of noninterviews verified in reinterview showed that several sample household units were misclassified and data falsification was detected. Although an entirely CATI reinterview program would be cost effective, several factors must seriously be considered to ensure that data is collected on all respondents and interview types.

Survey Satisficing in a Mixed Mode Military Survey

Yonghe Michael Yang, ICF International (myang@icfi.com); Amy E. Falcone, ICF International (afalcone@icfi.com); Lynn M. Milan, United States Army Research Institute (lynn.milan@us.army.mil)

Several studies have investigated the relationship between survey mode and data quality. They have shown that mode of data collection may influence a variety of survey errors, including coverage, sampling, measurement, and data processing errors. This paper reports our investigation into the effect of survey mode on one particular type of measurement error—survey satisficing—based on a mixed mode survey of the military population. Survey satisficing refers to the respondent's tendency to shortcut or skip the necessary cognitive process in answering survey questions. Existing theory of survey satisficing is based on the assumption that optimal question answering involves a series of cognitive steps. A respondent must interpret the meaning and intent of each question, retrieve relevant information from memory, integrate that information into a summary judgment, and report that judgment accurately. Some respondents may be unable or unwilling to carry out the required cognitive steps due to a lack of ability and/or motivation, task difficulty, or fatigue. In addition to respondent's individual characteristics, survey mode and question format have been found to influence the likelihood of satisficing in general population surveys.

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) adopted a mixed mode approach involving paper and Web-based data collections for its Sample Survey of Military Personnel (SSMP). This research focuses on the effect of survey mode as well as the interaction of survey mode with demographic variables on the likelihood of satisficing among the military population. We use three measures of satisficing: (1) non-differentiation—the tendency to provide identical or nearly identical responses to a battery of related questions using the same response scale; (2) no-opinion responding—the tendency to choose “don't know” or “not applicable” when such options are explicitly offered; and (3) acquiescence—the tendency to agree with any assertion regardless of its content.

Differences in Reliability and Validity Across Mail- And Phone-Administered Surveys

Sonja Ziniel, Harvard Medical School; Children's Hospital Boston (sonja.ziniel@childrens.harvard.edu); Nina Rauscher, Children's Hospital Boston (nina.rauscher@childrens.harvard.edu); Jennifer Koch, Children's Hospital Boston (jennifer.koch@childrens.harvard.edu); David Thompson, Children's Hospital Boston (david.thompson@childrens.harvard.edu)

Research on mode effects has increased over the past decades due to the popularity of mixed-mode survey designs. While a large number of studies have assessed potential factors for different response distributions across data collection modes, only a small number of studies addressed the influence of data collection mode on the psychometric properties of measurements, reliability and validity. Generally,

survey administration via mail compared to interviewer-administered survey seems to get more accurate answers to sensitive questions as well as higher internal consistency scores. No previous study, however, has been designed in a way that allows examination of differences in reliability as well as validity across different data collection modes taking into account characteristics of the survey administration, the time between measurements and question characteristics.

The Program for Patient Safety and Quality at the Children's Hospital Boston/Harvard Medical School has reviewed their nationally used Inpatient Experience Survey and designed a study to assess the psychometric properties of the about 100 measurements included in the new survey. Close to 475 people were recruited and 367 of them participated in all parts of the study. The data collection mode, phone or mail, was randomly assigned to the respondents after recruitment. The study design allows assessing differences in test-retest reliability and internal consistency between phone and mail modes. In addition, we can compare differences as a function of respondent and question characteristics and the time between the measurements. In addition, for a subset of the variables we are able to examine differences between modes in validity using medical record data as a gold standard.

How Do Respondents React When Asked to Self-Report Their Behavior?

Marla D. Cralley, Arbitron Inc. (marla.cralley@arbitron.com); Z. Tuba Suzer-Gurtekin, University of Michigan, Program in Survey Methodology (tsuzer@umich.edu)

Often self-report survey instruments such as time-use or diet diaries are used to measure specific respondent behaviors. Concerns arise around the accuracy of the reports. Do respondents accurately report their behavior, do they embellish their reports in an attempt to satisfy the researcher, and/or do they actually alter their behavior during the survey period?

Arbitron PPM, a system to passively collect Radio and Television media use over time among an on-going panel of respondents, has been implemented in 25 top U.S. metros, replacing the traditional paper Radio and Television self-report diaries previously used in these markets. During the Summer of 2008, Arbitron asked a test sample of PPM panelists to keep a one-week Radio diary while continuing to wear their PPM Television and Radio meters.

This paper will present the results of that study, comparisons of Radio listening passively collected through the PPM technology with respondent self-reports for the same time period. Patterns of differences will be compared and contrasted by demographic characteristics. In addition, passively collected listening information from diarykeeping panelists will be trended across several weeks to explore differences in actual behavior correlated with the diarykeeping task.

Assessing Data Quality across Probability Samples: An Examination of a Post General Election Mixed-Mode (Internet and Mail) and Telephone Survey

Lonna Rae Atkeson, University of New Mexico (atkeson@unm.edu); R. Michael Alvarez, California Institute of Technology (rma@hss.caltech.edu); Alex N. Adams, University of New Mexico (alexnadams@yahoo.com)

The Internet is a valuable survey tool and increasing in popularity. However, its use is still in its infancy and questions about survey response rates, representativeness and the effect of survey mode need to be addressed. After the 2008 election we were in a unique position of having two almost identical surveys that sampled from the same statewide voter registration file to collect data on New Mexico voters. One was a traditional telephone survey and the other a mixed-mode survey. Mixed-mode respondents were contacted three times via US mail and asked to complete our survey online or request a mail survey (80%

of respondents chose the Internet). Because of the nature of our sample population we know some sample demographic and election characteristics including gender, age, partisan registration, registration date, location, voting mode, voting history and several election outcomes. This allows us to compare the telephone and mixed-mode sample respondents to the sample population to assess their general representativeness. In addition, we also examine the influence of survey mode on response patterns. Because differences between survey results could be due to coverage differences or interview mode we use propensity score matching techniques to match telephone respondents with mixed-mode respondents on a set of covariates. This controls for coverage errors between the surveys allowing us to assess differences between survey modes. For example, we know that telephone respondents had greater voter confidence and were generally more optimistic than mixed-mode respondents. However, only by matching respondents across modes can we determine the magnitude of mode effects. This study sheds light on the use of mixed-mode and telephone surveys and their ability to create representative samples, even with relatively low response rates, the extent of coverage bias, and the influence of mode on survey response.

Opportunity Through Diversity

The 2008 National Asian American Survey: Context and the Civic and Political Engagement of Asian Americans.

Taeku Lee, University of California at Berkeley (taekulee@berkeley.edu); Jane Junn, University of Southern California (junn@usc.edu); Karthick Ramakrishnan, University of California, Riverside (karthick.ramakrishnan@ucr.edu); Janelle Wong, University of Southern California (janellew@usc.edu)

We present our findings from the 2008 National Asian American Survey (NAAS). The NAAS is the most comprehensive survey to date of the civic and political life of Asians in the United States. Based on 5,159 interviews conducted in August and September, 2008, the NAAS interviewed respondents from six primary Asian ethnic groups (Asian Indians, Chinese, Filipinos, Japanese, Koreans, and Vietnamese) and in eight languages (English, Cantonese, Mandarin, Korean, Vietnamese, Tagalog, Japanese, and Hindi). The NAAS also drew an innovative sample designed to be both nationally representative and include a sufficient number of cases to separately analyze Asian Americans in specific regional and metropolitan contexts (see www.naasurvey.com).

Our paper will focus on the determinants of Asian American political engagement. We define political engagement broadly to include voting in the context of the 2008 presidential election and as well as non-electoral and transnational political activities. We then explain these patterns of political engagement in terms of:

- Institutional-level factors such as the role of political parties, mass media, and religious and civic organizations.
- Individual-level factors such as sociodemographic background, immigrant socialization, political orientation, perceptions of racial in-group and out-group commonality, and experiences of discrimination.
- Contextual-level factors such as the demographic composition and economic conditions in one's zipcode, descriptive electoral representation at local, state, and federal levels, and state and local governance structures.

Our paper also pays special attention to ethnic and regional differences in defining the contours of “Asian America” in our survey respondents. An emerging context of particular interest is “new immigrant destinations” like Dallas/Fort Worth, Charlotte, Las Vegas, Salt Lake City, Minneapolis/St. Paul, and so on. We conclude by tying our findings to key issues confronting the Asian American community – e.g., the economy, immigration policy, health care, abortion, discrimination, hate crimes, language access.

The Global Survey of Physicists: Challenges and Lessons Learned from an evolving Multilingual Multinational Survey Effort.

Casey Langer Tesfaye, American Institute of Physics (clanger@aip.org); Rachel Ivie, American Institute of Physics (rivie@aip.org)

The Global Survey of Physicists is an evolving online survey effort produced by the American Institute of Physics (AIP). The survey applies nontraditional distribution methods to collect data that would otherwise be impossible to gather. The survey has been shaped and redesigned as a result of its challenges and successes. In 2002 the first global survey of women was a snowball survey that reached over 1000 respondents in 50 countries via e-mail and broke new ground in international data collection within the physics community. In 2005, a revised, improved and more standardized web-based global survey of women reached over 1300 respondents in 71 countries. In 2009, the web-based survey was expanded to an 8 language, 23 page web survey that has already dramatically outpaced the 2004 survey, collecting nearly eight times the number of responses to date (the survey is still open and collecting responses), including respondents from more than 150 countries. Although the current data collection effort is much more standardized than past efforts, including samples and censuses of physical societies in more than 14 countries in addition to more casual distribution efforts in 89 countries, the reach of the survey would have been deeply compromised if it was restricted to a representative sample.

This paper will address the evolving methodologies and challenges of the global survey, as well as its increasing effectiveness and reach. We will also examine the many complications associated with the collection of reliable and valid data in a small-scale multicultural context. Linguistic and cultural diversity are challenges that have amplified effects within the quantitatively driven, sensitive field of survey research, but are a very necessary part of the future of the field, as the globe becomes increasingly open, porous and permeable.

Opportunities for Contact with Culturally Diverse Populations: Interpersonal Contact and Immigration Attitudes.

Chase Harrison, Harvard Business School (charrison@hbs.edu)

In recent years, the ethnic and cultural diversity of the United States has increased dramatically, changing both the number of opportunities for interactions between those of different ethnic groups, as well as the diversity of the types of contacts. This paper will study the impact of inter-personal contact with immigrants on attitudes toward public opinion about immigration policy as well as attitudes about cultural diversity in the US.

Opposition to immigration has typically been observed in two dimensions: cultural and material. Cultural or symbolic attitudes toward immigration typically have centered on differences in language, ethnicity, and customs between new immigrants compared to more established groups. Material concerns about immigration have particularly focused on job competition between native-born Americans and recent immigrants.

Residential contexts and personal contacts have been used to explain the formation of attitudes toward racial and ethnic groups. Proximity has been found to heighten perceived resource-competition between groups who live in proximity to one another. On the other hand, direct interpersonal contact between members of different groups has been understood to reduce animosity and prejudice. This paper analyzes the impact that inter-personal contact within different settings has on attitudes about immigration and immigrants while controlling for residential proximity.

Data will be drawn from two Random-Digit-Dial (RDD) surveys of American adults conducted in 2001 and 2002. Multivariate models, including relevant controls, will be estimated among 1,348 cases pooled across the two cross-sectional surveys. Residential context is measured using U.S. Census data and derived estimates, while inter-personal contact is based on self-report.

Representing Hispanics in General Public Opinion Surveys.

Jocelyn Kiley, Pew Research Center (jkiley@pewresearch.org); Leah Christian, Pew Research Center (lchristian@pewresearch.org); Michael Dimock, Pew Research Center (mdimock@pewresearch.org); Scott Keeter, Pew Research Center (skeeter@pewresearch.org)

As the U.S. population continues to grow more ethnically diverse, many public opinion surveys conducted in English do not allow for reliable reporting on the opinions of Hispanics because of concerns about both representativeness and inadequate sample sizes. Yet, due to time and financial constraints, bilingual interviewing – that is, fully bilingual interviewers and translation of questionnaires before interviewing begins – which addresses both issues, is often not a viable option for practitioners.

In 2009, the Pew Research Center for the People & the Press conducted several surveys—including one with a large Hispanic oversample—that employed a “middle ground” approach in which all initial contacts were made in English, and households screened for language were called back by Spanish-speaking interviewers. In this protocol, questionnaires could be translated during the first night of the field period. The surveys included questions about country of origin, immigrant generation, and language proficiency. This allows for a detailed comparison of the Hispanic sample in these surveys with estimates of the Hispanic population from the American Community Survey as well as with surveys that utilize advance translations and fully bilingual interviewers. Analysis to date finds that the middle ground approach may under represent foreign-born Hispanics, those of Mexican origin, and those living in the Pacific West. Additional analysis for this paper will examine the impact of these generational and regional differences on opinion through comparisons with substantive questions on fully bilingual Pew Hispanic Center surveys.

The paper will also address the issue of how best to identify Hispanic respondents: whether to ask if a person is ‘Hispanic’ or “of Hispanic origin or descent.” Analysis of several question wording experiments in 2009 suggests that variations in how Hispanic ethnicity is asked do not have a significant effect on the percentage of Hispanics identified.

Questionnaire Design: Structure, Response Style, and Self-Reports

Effect of Questionnaire Structure on Nonresponse and Measurement Error: Sequential vs. Grouped Placement of Filter Questions

Lisa R Carley-Baxter, RTI International (lcbaxter@rti.org); Andy Peytchev, RTI International (apeytchev@rti.org); Michele Lynberg Black, CDC (mc12@cdc.gov)

Based on a cognitive perspective, survey practitioners aim to structure questionnaires in the way that respondents' memories are organized. If details about multiple behaviors are requested, asking the respondent to report a type of behavior and then following up with more detailed questions is assumed to help recall, reduce measurement error and possibly decrease item nonresponse.

When surveys ask about multiple behaviors, a fear arises that asking many detailed questions after such a filter question will deter some respondents from reporting the next type of behavior on a filter question, inducing measurement error bias in survey estimates. This may also lead to higher breakoff rates prior to answering all key filter questions and higher item nonresponse rates to the follow-up questions.

There is some evidence from self-administered surveys that when respondents recognize such a skip pattern, they become more likely to provide responses that avoid subsequent questions. The lack of interviewer presence may encourage this behavior. There is also similar limited evidence on grouping filter questions from in-person interviewing, yet coming from a study with a relatively long interview. Furthermore, the high interviewer rapport in face-to-face interviews may mask potential effects on breakoff and item nonresponse.

We randomly assigned respondents in a dual-frame landline-cell phone study on intimate partner violence to either a sequential or grouped filter question design. In both versions, respondents were asked to identify the number of perpetrators for different types of victimizations. In the sequential design, multiple in-depth follow-up questions about each reported person were asked immediately after each filter question. The study is being conducted in November and December, 2009. Outcome measures include the number of reported perpetrators, item nonresponse, breakoff, and interviewer feedback.

Response Scales' Vulnerability to Acquiescence and Extreme Response Style Behavior
Natalia Kieruj, Department of Methodology and Statistics, Tilburg University (n.d.kieruj@uvt.nl);
Guy Moors, Department of Methodology and Statistics, Tilburg University (Guy.Moors@uvt.nl)

Acquiescence – the tendency to agree with opinion questions regardless of the content – and extreme response style behavior – the tendency to pick the extremes of a response scale while avoiding intermediate response options – may distort the measurement of attitudes. How response bias is evoked is still subject of research. A key question is whether it may be evoked by external factors (e.g. test conditions or fatigue) or whether it can be the result of internal factors (e.g. personal answering styles or socio-demographic factors). If the former is the case, it is implied that creating optimal test conditions can reduce the occurrence of these biases. However, if the latter is the case, response bias is difficult to prevent from happening and therefore should somehow be corrected for.

In the first part of this study we explore whether scale length – which is the manipulated test condition – influences the occurrence of extreme response behavior and/or acquiescence, i.e. by varying from 5- till 11-point scales. In pursue of this we apply a latent-class D-factor model that allows for diagnosing and correcting for ERS and ARS simultaneously. Results show that scale length has little effect on ERS since it appears to be a characteristic that returns in each of the treatments. Only weak evidence of ARS is found.

In a second study we check if ERS might reflect a internal personal style by (a) linking it to external measures of ERS, and by (b) correlating it with particular personality characteristics (e.g. extraversion and indifference). Results show that ERS is reasonably stable over questionnaires and that it is associated with the selected personality characteristics.

Making It All Add Up: A Comparison of Constant Sum Tasks on Self-reported Behavior

Ronald Z. Szoc, ICF International (rszoc@icfi.com); Randall K. Thomas, ICF International (randall.k.thomas@gmail.com); Frances M. Barlas, ICF International (fbarlas@icfi.com)

Constant sum tasks help determine the relative importance of elements by giving respondents a fixed number of units of some resource, such as points, dollars, or percentages, to allocate among those elements. In market research, the task may consist of allocating money for different consumer goods within a particular class of products based on the overall amount consumers say they spend for that class of products (a 'share-of-wallet' exercise). We examined three different constant sum tasks and 2 alternative response formats in a web-based survey. Respondents were randomly assigned to one of two topics and once assigned to topic, respondents were randomly assigned to evaluate 8 elements with one of 10 response formats: five involved dollar allocation and five involved unit allocation. While the order of allocation for the elements was comparable across formats (units and percentages were in similar order for the elements), constant sum measures generally had lower criterion-related validity than the alternative formats with two different criteria we examined. Differences in validity coefficients were large, though we will show how various transformations of the predictors and criteria could significantly alter conclusions.

Function Follows Form: Response Format Effects on Self-reported Individual and Household Disability

Amy E. Falcone, ICF International (afalcone@icfi.com); Randall K. Thomas, ICF International (randall.k.thomas@gmail.com)

Disabilities have been conceptualized and measured differently in national surveys, both at individual and household levels. The prevalence of disabilities as established by self-report surveys often affects policy and resource allocation decisions and requires reliable and valid measurement. This study evaluated 3 different response formats on the reporting of disability for both self and other household members (for visual/auditory/cognitive/physical impairments). In a web-based survey, respondents were randomly assigned to one of three groups: 1) presented disability items separately for self and household members using YNG; 2) presented disability items separately for self and household using MRF; or 3) presented disability items and asked to select whether self, another household member, self and another household member, or no one in household had each impairment (Combination Grid – CG). Though the CG format took the least amount of time to complete overall (and YNG=MRF in time), the YNG and CG formats generally resulted in similar and higher prevalence levels of most disabilities for both self and household levels than MRF. We discuss a number of results that indicated that disability salience may be a more important determinant of response selection with MRF than other response formats but MRF may incompletely describe prevalence.

AAPOR Poster Session 3

Saturday, May 15, 2010

1:15 p.m. - 2:15 p.m.

Poster Session - 3

Assessing Online Media Use and Digital Interpersonal Communication Variables as Predictors of Political Knowledge

Jason A. Martin, Indiana University School of Journalism (jam15@indiana.edu)

Online media played a substantially larger role in the 2008 presidential campaign than ever before, including innovations in how candidates contacted potential voters and how citizens discussed public affairs. This paper examines how developments in online and mobile media technologies are changing the ways people consume and exchange news, gain knowledge, and form opinions.

Using nationally representative data from the 2008 Pew Center for the People and the Press Biennial Media Consumption Survey, secondary analysis was performed to examine how online measures of media use and political discussion might apply to the traditional theoretical model of education, media use, and interpersonal communication as predictors of political knowledge.

Results indicate that online media use has emerged as a significant predictor of knowledge independent of traditional media use. Also, mobile device use was significantly correlated with knowledge for several different measures.

While some findings supported the idea of the democratizing potential of the Internet, mixed results raise the possibility that increased media choice actually expands the knowledge gap. Respondents from states that scored better on an Internet infrastructure and business climate index were more likely to have higher levels of knowledge than states with less online capability and online business activity.

Additionally, regression analysis found online news use and interpersonal communication, as measured through news stories exchanged through e-mails, to be statistically significant and methodologically important predictors of knowledge when controlling for traditional media use and demographics.

These findings indicate that online media use and interpersonal communication variables should receive further consideration in public opinion research. With use of online public affairs news ever-increasing, these results contribute to a greater understanding of how people access and discuss news online, how researchers might measure those exchanges, and how those habits may translate into political knowledge and opinion formation.

On the Usefulness of Pretesting Vignettes for Exploratory Research

Jennifer Beck, U.S. Census Bureau (jennifer.l.beck@census.gov)

Survey methodologists use vignettes as an evaluative tool for pretesting survey questionnaires. However, these fictional scenarios also lend themselves to exploratory, substantive research about the topics upon which surveys are based. Researchers across multiple disciplines have used vignettes to determine how

people make judgments and decisions across a variety of complex and potentially sensitive situations. In this paper, I discuss the psychological framework for why vignettes are a useful tool for uncovering people's judgments and decision making processes. As an example of how to use vignettes as an exploratory tool, I interpret findings from pretesting vignettes in light of what they reveal about judgment and decision making. More specifically, I present evidence that preliminarily reveals information about how kids and teens think about activities that are relevant to their vulnerability to Internet predators. These findings suggest several key variables for researchers to explore in an effort to improve education about internet crimes against children.

Use of Surveys and Cluster Analysis to Increase Ridership on Mass Transit: A Case Study
Frank Lynch, Abt SRBI (f.lynch@srbi.com); Jason Minser, Abt SRBI (formerly with the Chicago Transit Authority) (j.minser@srbi.com); Mindy Rhindress, Abt SRBI (m.rhindress@srbi.com)

With Americans grappling with high energy prices and worries over greenhouse-gas emitting fossil fuels, this survey represents a major research effort to find actionable ways to increase public transit usage. This poster session will demonstrate the use of surveys and special analytic techniques to inform public agencies and promote mass transit usage. The poster session will unveil specific conclusions and recommendations based upon the survey analysis.

Specifically, this project, conducted for the Chicago Transit Authority (CTA), one of the nation's largest transit agencies, is based on data from a 2009 survey of 2,800 area residents, conducted by telephone.

The survey analysis uses procedures such as K-means clustering techniques to identify clusters of individuals who share similarities but are different from those in other groups. The analysis leads to identifying areas for improvement for the CTA, as well as the development of communications strategies to promote mass transit ridership in Chicago. Key inputs to the analysis include the identification of clusters within the market as a whole, perceptions of CTA strengths and weaknesses, and perceived comparative advantages and disadvantages vs. private vehicles.

The Chicago Transit Authority (CTA) provides both bus and subway (EL) transit service in the city and many Cook County suburbs. In 2008 the CTA exceeded a half billion riders.

Sample Disposition of RDD Unlisted, RDD Listed, and Cell Phone Only Telephone Households
John Tarnai, Social & Economic Sciences Research Center (tarnai@wsu.edu); Danna L Moore, Social & Economic Sciences Research Center (moored@wsu.edu); Marion Schultz, Social & Economic Sciences Research Center (Schultzm@wsu.edu)

Telephone survey samples of the general public may use combinations of three different sample frames including random digit dialing (RDD), listed households, and cell phone only households, to adequately represent all types of households in a survey. A previous paper (Tarnai, Moore, Schultz, 2009) showed that there are significant demographic differences among respondents from these sample frames. The present paper explores the sample disposition differences among these three sample frames, from eight similar surveys conducted over a two year period. Specifically, the paper examines the rates of several standard AAPOR sample disposition categories, including (a) non-working telephone numbers; (b) non-contactable households; (c) ineligible households; (d) refusals; and (e) partial completed interviews. Each of the eight surveys included in this paper include all three sample frames, and starting sample sizes of up to 8,000 numbers, and a minimum of 400 completed CATI interviews. The analysis focuses on the differences in these sample disposition rates by sample frame, and over time. Preliminary results indicate substantial differences in rates by sample frame, as well as changes over time in some of these rates.

This information is useful for survey researchers who are planning telephone surveys that use these sample frames, to accurately estimate these sample disposition rates.

Experimental Trial of Benefit Appeals on Completion Rates for the Agricultural Screening Survey

Danna L. Moore, Social and Economic Sciences Research Center (moored@wsu.edu); Dale Atkinson, USDA/NASS/RDD (Dale_Atkinson@nass.usda.gov); Jaki S. McCarthy, USDA/NASS (Jaki_McCarthy@nass.usda.gov)

This study empirically examines the effects of two different prominently displayed appeals in combination with set confidentiality assurances and other survey statements on completion rates for agricultural screening questionnaires. The experiment was carried out on a sample of 13,000 farm units, using person name and business entity (farm name) addressed pre-notification letters and mail back questionnaires. While web surveys have become increasingly more common in research, this methodology has not been thoroughly investigated for agricultural populations. Whether web can be an effective methodology for reducing more expensive in-person interviewing and land visits for determining eligibility for the Agricultural Census is evaluated. A critical question in this research was whether a pre-notification letter asking respondents to complete a short 5 minute web screening questionnaire about their agricultural involvement could be effective. Also of interest was whether benefit appeal differences would hold across mixed mode (web, mail, and telephone) administration and whether systematic differences would contribute to coverage differences and response effects.

Survey Weight Adjustment using PROC WTADJUST from SUDAAN V10.

G Gordon Brown, RTI International (ggbrown@rti.org)

Analysis weights are used in surveys to reduce bias and improve inference. Creating analysis weights that take into account survey design is a common pre-analysis task for any survey. Frequently, this task can be difficult and time consuming and requires the input of a weighting specialist. This paper will demonstrate how analysis weights can be created using the new WTADJUST procedure that is currently available in the software package SUDAAN® V10. This procedure is easy to use and is particularly useful for small to mid-sized surveys.

The WTADJUST procedure allows the user to compute sample weight adjustments to account for features such as non-response and coverage error (post-stratification) using a model based calibration approach. This approach is based on a logistic model and allows the user a great deal of flexibility when creating analysis weights. Other features of the WTADJUST procedure include weight centering, trimming of extreme weights, and subsetting.

I will present one complete example where the WTADJUST procedure was used and several shorter examples. The complete example will demonstrate how WTADJUST was used to create the analysis weights for the Maryland Dental Survey. This small survey attempted to contact persons in the state of Maryland that had visited a dentist, emergency room, or surgeon to obtain emergency dental services. Data collected from the survey included questions about the quality of care received, cost of care received, and the attitude of staff. This talk will demonstrate how post-stratification and weight trimming were used to create the final analysis weights that were used in the study. Shorter examples will demonstrate non-response adjustment and creating analysis weights for sub-domains of interest.

Online Table Generator Tool for Health Insurance Coverage Estimates

Karen Soderberg, SHADAC - University of Minnesota (soder145@umn.edu); Michael Davern, NORC (davern-michael@norc.org); Jeanette Ziegenfuss, Mayo Clinic

(Ziegenfuss.Jeanette@mayo.edu); Peter Graven, SHADAC - University of Minnesota (grave165@umn.edu); Michele Burlew, SHADAC - University of Minnesota (mmburlew@umn.edu)

Software and computing tools allow for increasingly fast analysis of complex survey data. However, these tools require funds to purchase the software and training to properly apply procedures. Because of these barriers the access to data not necessarily expanded for many users, particularly underfunded researchers and policy analysts that do not have specialized staff. The State Health Access Data Assistance Center (SHADAC) identified this as a barrier for several of its audiences and has developed an online table generator tool that allows easy access to estimates of health insurance coverage. This poster will describe how data warehousing technology was utilized to provide a flexible and user-friendly set of estimates, including counts, percents and standard errors. Data sources include the Current Population Survey's Annual Social and Economic Supplement from 1988 to present with single, two-year and three-year averages, and the first-ever health insurance coverage estimates from the American Community Survey. The poster will include an overview of the tool, the development process, the content management system, promotional efforts, and planned enhancements.

Multiple Identities: Effects of Question Framing on Self-rated Importance of Factors Influencing Self-concept

Frances M. Barlas, ICF International (fbarlas@icfi.com); Randall K. Thomas, ICF International (randall.k.thomas@gmail.com); Ronald Z. Szoc, ICF International (rszoc@icfi.com)

The sense of self is formed by repeated interactions with others and the world around us, shaping our subsequent interactions and opinions about the world. Many define the sense of self as the collection of values, interests, and personality that uniquely define the individual, while others define identity more holistically as a unique sense of self apart from the constituent values, interests, and personality. Differences in conceptualization can lead to differences in measurement and results. This study examined whether people reacted differently when presented with different question contexts (i.e., asking about influences on their values versus influences on their identity). Respondents participated in a web-based survey and were asked to rate the importance of 16 items (including race, sex, occupation, spouse/partner, etc.) in either (randomly assigned) 'defining your values, interests, and personality' or 'defining your identity and who you are.' We found significant differences in how important the items were rated based on context. Correlations of the items were somewhat higher with the collectivistic-individualistic measures of cultural values with the 'values' rating context. In addition, political ideology appeared somewhat more related to the 'values' context than 'identity' context.

When to Email or Not?: Determining the Best Day of the Week to Send Evites For Participation in a Web Survey.

Sylvia R. Epps, Decision Information Resources, Inc. (sepps@dir-online.com); Leslyn Hall, Redstone Research, LLC (leslyn.hall@gmail.com); Jo Anna Hunter, MDRC (joanna.hunter@mdrc.org)

Over the last decade, there has been an increasing popularity in using Web surveys to study various populations. In particular, high usage of the Web among college students lends itself as a cost-effective data collection strategy for this population (Carini, Hayek, Kuh, Kennedy, & Ouimet, 2003). As a result, researchers have focused on whether web surveys deliver comparable data to paper surveys, finding that differences in data quality are usually minimal (Carini et al., 2003), though concerns about data security and nonresponse bias remain (Smith 1997). Still, given using Web surveys as the main mode of a data collection is relatively new, many important questions related to when and how often to contact respondents to ensure high response rates remain unanswered. For example, should email invitations

(evites) be sent on a particular day of the week; which day yields the highest same-day response? These questions form the basis of the proposed study.

Specifically, we will use data from 5 completed mix-mode data collection efforts of college students (Web-Telephone-In-person follow-up) to examine the “best” day to contact respondents (through evites) to obtain completed surveys. Across these 5 studies (with ~1668 Web completes), we will evaluate the following:

1. Sending evites to respondents on which day of the week will yield the greatest number of same-day responses?
2. Sending evites on which day of the week will result in the smallest average number of days between evites and survey completion?
3. Sending evites on which day of the week will result in completed surveys with higher data quality?

Initial results indicate that compared to later in the week, contacting students earlier in the week (Monday and Wednesday) results in a greater number of completed surveys, a shorter time to completion, and that there is no discernable effect on data quality.

Exploring the Use of Self-Reported Cell Phones in Voter Files for Likely Voter Sampling
Patrick Faust, Greenberg Quinlan Rosner Research (pfaust@gqrr.com); Adam Slater, Greenberg Quinlan Rosner Research (ASlater@gqrr.com)

Our research seeks to answer the question on whether or not cell phones reported in the voter file as main telephone contact can replace or supplement random digit cell phone sampling for a likely voter universe. Cell phone sampling is traditionally limited to large geographic units due to certain smaller geographies not coinciding with cell exchanges. Our approach supports sampling at this level on a case by case basis, if the voter file in the district has representative cell phone only records to support this type of sampling in a methodologically sound manner.

Extensive research has been conducted on how to solve the “cell phone-only problem” (Link et al, 2008). This work has focused on various dual sample frame approaches to solving the rising coverage issue with cell phone only households using Address Based Sampling (ABS). This approach works well for researchers with longer fielding windows but is not practical for political pollsters who need quick turnaround with more cost effective way of talking to these cell phone only voters (CPO).

Our research will use a combined dataset of RDD cell phone interviews from the monthly Democracy Corps tracking surveys in 2009. We seek to analyze any difference in political attitude between this universe and a national voter file derived from self reported cell phones existing in the national voter file, identified through commercial identification of known cell phone exchanges. Finally, our research will expand upon the potential implications on this type of sampling for political pollsters with respect to bias and coverage, cost efficiency, and the ability to conduct political survey at the congressional level with CPO households.

Offering a Web Option in a Mail Survey of Young Adults: Impact on Survey Quality
Scott Turner, Fors Marsh Group, LLC (sturner@forsmarshgroup.com); Luciano Viera Jr., Fors Marsh Group, LLC (lviera@forsmarshgroup.com); Sean Marsh, Fors Marsh Group, LLC (smarsh@forsmarshgroup.com)

The increased prevalence of Internet usage has prompted survey administrators to offer respondents a web option for submitting their survey responses. Collection of survey data via the web is particularly attractive as it is often more cost-effective when compared with paper-based, telephone-based, and in-person interview modes. The addition of a web option is also thought to improve survey quality by increasing response rates, particularly among young adults who may be more likely to expect such an option and would otherwise not respond to other survey modes. However, very little work has experimentally examined this hypothesis.

This paper will present the results of an experiment designed to test the impact of offering a web option on the survey quality of a mail survey tracking the future career plans of young adults. The survey administration process consisted of five mailings designed to maximize survey response rates. A total sample of 20,000 young adults 16-24 were randomly assigned to one of three conditions:

- 1) Mail only (10,000 cases) – were given the option of completing the survey via the paper questionnaire only.
- 2) Mail or Web (5,000 cases) – were given the choice of completing the survey via the paper or the web.
- 3) Web only (5,000 cases) – were given the option of completing the survey via the web only.

Specific measures of survey quality will include survey response rates, respondent profiles, and key metrics. Additionally, because mail surveys have been proposed as a means of capturing cell only households, respondent telephone usage will also be examined to determine if offering a web option has any additional benefit in reaching young adults who live in cell phone only households. Implications for existing survey practice and directions for future research will be discussed.

Are Estimates of the Medicaid Undercount Stable Over Time?

Robert Kirby Goidel, Louisiana State University (kgoidel@lsu.edu); Stephen Barnes, Louisiana State University (barnes@lsu.edu); Dek Terrell, Louisiana State University (mdterre@lsu.edu)

The Louisiana Health Insurance Survey is a biennial survey of 10,000 households designed to provide detailed estimates of uninsured children and adults. Since 2005, each survey has included a sample of over 1,000 Medicaid recipients as a mechanism for estimating the effect of the Medicaid undercount on uninsured rates. In this paper, we examine estimates of the Medicaid undercount in three different years (2005, 2007, and 2009). The results allow us to consider the stability of misreporting over time. Using these results we then develop a model to correct for misreporting based on the probability that a given individual misreported his/her insurance status. Finally, we apply the model to the 2008 American Community Survey estimates of uninsured adults and children and compare these results to the findings from the 2009 Louisiana Health Insurance Survey.

The Subject Lines of Web Survey Invitations and Participation Rates

Handan Titiz, Children's Hospital Boston (handan.titiz@childrens.harvard.edu); Sonja Ziniel, Children's Hospital Boston (sonja.ziniel@childrens.harvard.edu)

Web surveys are an efficient data collection mode if surveys are sent out within an organization. While the survey period is shorter compared to other modes and can provide results sooner, web surveys generally suffer from low response rates. Numerous studies have been conducted on increasing response rates in web surveys using incentives, reducing length, multiple contacts, and alike. However, the work on how the subject line of the invitation email affects the likelihood of participation is limited. One study reports that for groups less attached to the study sponsor, there was a significant difference in subjects' participation when different subject line contents were used. The respondents with high

attachment to the sponsor did not show any significant differences in participation across different subject lines (Porter & Whitcomb, 2005). Another study shows a difference in favor of email subject lines that are in a “plea” rather than an “offer” format (Trouteaud, 2004).

Survey fatigue has been an increasing concern for the field of survey research. Researchers at Children’s Hospital Boston/Harvard Medical School receive numerous web survey invitations per month. One such survey is the Children’s Hospital Boston’s Clinical Research Program annual service satisfaction survey. Using this survey we implemented a 2x2 experiment with regard to the subject line of the invitation email. We hypothesized that the use of the word “survey” itself in the subject line diminishes the likelihood of participation because of the participants’ survey fatigue. We also expected to replicate Trouteaud’s results with respect to the use of “plea” versus “offer” format in the invitation email’s subject line. The participants were randomly assigned to each of the four conditions. This study, therefore, allows us to examine if the content of the subject line of the invitation email to a web survey influences the likelihood to respond.

Nonresponse in a Dual Frame Telephone Study: Illinois Children’s Insurance Research

Julie A. Pacer, Abt SRBI (pacer@srbi.com); Kelly Daley, Abt SRBI (daley@srbi.com); Michael P. Battaglia, Abt Associates (Mike_Battaglia@abtassoc.com); Dianne Rucinski, University of Illinois-Chicago (drucin@uic.edu)

The considerable increase in the number of cell-only households in recent years has resulted in telephone surveys that draw sample from both landline and cell phone sample frames. While this solution addresses coverage error, it neglects nonresponse error. The effect of nonresponse error in a dual frame telephone sample is important because response propensity may be related to the sample frame (Brick, 2009).

Abt SRBI is currently conducting the Illinois Children’s Insurance Research Study in conjunction with the School of Public Health at the University of Illinois-Chicago, to assess the impact of a state sponsored health insurance program (AllKids) on reducing the number and percentage of uninsured children in the state of Illinois over the past decade. This study is being conducted by telephone and employs a random digit dial (RDD) landline sample and an RDD cell phone supplement. Because this study utilizes a dual frame landline and cell phone sample, an opportunity exists to identify and compare nonresponse bias indicators in this survey to nonresponse bias indicators in similar recent studies (Brick et al., 2006; Brick et al., 2007; Keeter et al., 2009). The proposed research seeks to examine the extent of nonresponse due to accessibility bias (landline mainly users versus cell-phone mainly users) resulting from the dual sample frame. The Illinois Children’s Insurance Research Study offers a unique perspective in that the unit of random selection is family within household rather than adult within household. The impact of this non-traditional selection unit will be explored as well.

Revision to Teacher Nonresponse Bias Analysis

Danielle Kay-Maxwell Battle, American Institutes for Research (dbattle@air.org); Jared Coopersmith, Quality Information Partners, Inc. (ctr_jcoopersmith@air.org)

To examine different methods of nonresponse bias the authors tested a revised nonresponse bias analysis using the 2007-08 Schools and Staffing Survey (SASS), sponsored by the National Center for Education Statistics (NCES), and compared the outcomes to the previous method used for the survey. The new analysis gave direct, quantifiable measurements of bias between respondents and the sample population using frame characteristics. Chi-square tests were used to compare the distribution of the characteristics between the respondent and sample populations. The previous method used a set of

criteria to determine whether or not differences between the respondents and the sample frame were significant, but did not produce a quantifiable measure of bias. The authors found the new method yielded similar results for nonresponse bias as the previous method but offered several improvements. The quantified measurements of bias allowed direct comparison between different frame characteristics, as well as the ability to summarize bias levels across the various frame characteristics utilized. The new method also facilitated comparisons of bias before and after nonresponse adjustments and could be completed in less time with greater efficiency.

Does Launching a More Expensive Mode of Data Collection Earlier Save You Money?

Sylvia R. Epps, Decision Information Resources, Inc. (sepps@dir-online.com); Pam V. Wells, Decision Information Resources, Inc. (pwells@dir-online.com); Jo Anna Hunter, MDRC (joanna.hunter@mdrc.org)

In an era of declining response rates for all modes of survey data collection, mixed-mode designs are often adopted as a strategy to maximize response rates while also controlling survey costs (de Leeuw, 2005). Single mode designs, in general, can be cheaper to administer and easier to implement than a mixed-mode approach, and yet response rates often lag due to respondents' preferences and coverage issues. In the proposed study, we use data from two completed mixed-mode studies (Web-CATI-In-person follow-up) involving college students. Each study began with a Web-only period, followed by a CATI follow-up with non-responders. We ended the fielding period with an in-person follow-up with the remaining non-responders. We contacted respondents using email invitations and a reminder postcard during the entire duration of the study.

Typically, researchers tend to extend the least the expensive data collection mode to control costs; however, we have found that launching the more intensive and expensive data collection strategies sooner can ultimately result in cost savings while still achieving similarly high response rates. We will present data on the fixed costs of completing ~1830 surveys (which spans three cohorts of data collection) across the three modes of data collection as a whole, and then separately by mode to address the following:

- Does launching more expensive data collection modes (CATI and In-person follow up) earlier result in a lower fixed cost per complete, as compared to extending less expensive data collection modes (Web)?

Preliminary analyses indicate that there are savings associated with launching the more expensive data collection modes earlier (CATI and In-person follow up). Ultimately, the findings will improve our knowledge around data collection planning and assist other researchers when determining which mode to launch and when.

Imputation using SUDAAN PROC HOTDECK for Education Surveys

Darryl Creel, RTI International (dcreel@rti.org)

Most surveys implement methods to reduce nonresponse, unfortunately almost all surveys experience some level of nonresponse at the unit and item level. In this paper, we focus on the potential impact of item nonresponse on survey estimates and, primarily, on using the new HOTDECK procedure from SUDAAN® 10.0 to address item nonresponse. One method of analyzing data with item nonresponse is complete case analysis. That is, only analyze the data that do not have any missing values related to the analysis. One problem with complete case analysis is that it ignores any information that the item nonrespondents may have which creates the potential for biased survey estimates, if the missingness is not missing completely at random.

An alternative to complete case analysis is to fill in or impute the missing items. Analyzing data with imputed values has the advantage of keeping all of the data available for analysis and, hopefully, minimizing the potential for biased survey estimates. Although imputation can be a resource intensive task, one way to reduce the resources allocated to the imputation task is to use flexible standardized software. SUDAAN® 10.0 has introduced a procedure called PROC HOTDECK which implements the weighted sequential hot deck imputation methodology. We demonstrate the application of PROC HOTDECK to impute a single variable on an education survey, to simultaneously impute multiple variables on an education survey, and to multiply impute through Monte Carlo simulation.

Ideology, Polarization and American Media Consumption: An Examination of Public Opinion Formulation

Anthony DiMaggio, Illinois State University (adimagg@ilstu.edu)

This paper examines the interplay between media consumption and opinion formulation on public policy related issues. It theorizes specifically about partisan media sources, including cable news outlets such as MSNBC and Fox News, the editorial pages at the New York Times and Wall Street Journal, and conservative radio.

Statistical analysis in this paper suggests that, after controlling for relevant demographic variables, media consumption plays an important role in influencing not only what consumers think about, but what they specifically think regarding various political issues. In short, partisan media consumption appears to play an important role in polarizing viewers in favor of more conservative and liberal views.

Media polarization, however, is not a one-way process. My analysis also finds that increased consumption of partisan media (whether conservative or liberal) is accompanied by increased openness to competing media outlets. Consumers of Fox News, for example, are more likely to also watch programs on MSNBC and other liberally oriented sources than those who do not watch Fox News. My review also finds that those who consume partisan sources also tend to perform quite well on civics I.Q. tests, suggesting that the allegedly negative effects of partisan news consumption may be exaggerated. Polarized news consumers, it appears, are quite actively engaged in the political process.

Favorable Results from a Low Effort Follow-up: An Investigation of Nonresponse Bias in a Customer Satisfaction Survey.

Jennifer E. O'Brien, Westat (obrienj@westat.com); Tom Krenzke, Westat (TomKrenzke@westat.com); David Ferraro, Westat (DavidFerraro@westat.com); Howard King, Westat (HowardKing@westat.com); Michele Harmon, Westat (MicheleHarmon@westat.com); Jocelyn Newsome, Westat (JocelynNewsome@westat.com); Kerry Levin, Westat (KerryLevin@westat.com); Martin Rater, United States Patent and Trademark Office (Martin.Rater@USPTO.GOV)

Under contract with the United States Patent and Trademark Office (USPTO), Westat administers the Customer Quality Panel Survey (CPQS), a longitudinal, overlapping panel customer survey. The CPQS is a web/mail survey conducted in quarterly waves with customers (e.g., agents, attorneys, sole practitioners) affiliated with the USPTO's top-filing firms. Top filing firms are those that have filed six or more patent applications in the past year and are more likely to be aware of day-to-day changes occurring within the USPTO. We were interested in whether responders and nonresponders differed in their perceptions of patent examination quality. Following the end of normal data collection, a postcard

with a single survey item printed on it was sent to nonresponders that were rotating out of the sample. The objective was to get a general understanding of how the nonresponders might have responded to the survey had they actually participated. A single key survey item ('Consider your experiences with USPTO Patent Examiners in the past three months. How would you rate overall examination quality for this time period?') was used to assess this potential bias. A chi-square test showed a significant relationship between the responses to the item and whether or not the respondent came from the quarterly wave sample or the followup sample. The responses from nonresponders allowed us to estimate the amount and direction of nonresponse bias. Specifically, follow-up responses indicated more favorable assessments of the quality of patent examinations. To help validate the results, the same process was conducted during the subsequent wave of data collection with nearly identical results. Thus, while there is likely potential for bias in the followup sample estimates, the results are an indication that perceptions of overall examination quality are more favorable than is indicated by considering the responses of customers in the quarterly wave sample alone.

Different Topics, Different Results.

Gregory A. Smith, Pew Research Center (gsmith@pewforum.org)

In May 2009, however, the Pew Research Center conducted a new survey, the primary focus of which was not religion, but science (scientific knowledge, views of scientists and attitudes about scientific controversies). The new science survey included the same question about human origins and development that had previously been asked in religion surveys. But in contrast with previous Pew polling, the science survey unexpectedly found a higher-than-normal belief in Darwinian evolution and fewer people expressing a creationist point of view.

This paper examines the reasons for this large and potentially substantively important discrepancy. Can the differences between surveys be attributed to differences in sample characteristics? Are there particular religious and demographic groups among whom views differ most markedly? Can the differences between surveys be attributed to differences in question context? And, perhaps most importantly, if all of these questions are answered negatively, does that suggest that topic salience bias may be an underlying cause of concern for this question and questions like it?

Comparison of Influences on Seat Belt Use in States

Cynthia Augustine, RTI International (caugustine@rti.org)

This paper models factors known to influence seat belt use across states, as well as factors involving data collection methodology, and discusses how these factors might influence future highway safety survey protocols. Of particular interest to the author is whether the data collection mode or interviewer characteristics influence driver and front passenger seat belt use rates. We are also interested in determining if factors that influence seat belt use are consistent between states, and if the populations in different states require different strategies to increase compliance.

Currently, some states are experiencing diminishing returns on seat belt enforcement campaigns like the Click-it-or-Ticket program. The results of this analysis will help to identify best practices in seat belt enforcement that can improve compliance across states, reduce the cost of data collection, programming and enforcement, and potentially save lives. Results may indicate the need to implement programs that target specific age groups or point toward state-level legislative action that improves compliance.

This research uses a dataset constructed from publicly available information about seat belt use rates in all 50 states. The dataset will include data collection characteristics and safety rate information as well as state legislative, enforcement and demographic information. Data collection information includes

the number of sampled sites, mode of data collection and interviewer characteristics. The type of seat belt law, the size of seat belt fines and the presence of driver license points are examples of the legislative data. Enforcement information includes the presence of high-visibility safety campaigns such as Click-It-Or-Ticket, adopted by most states, as well as information on the number of traffic fatalities. Demographic information includes state characteristics such as urbanicity, roadway types and diversity of population.

A Reasonable Use of Non-Probability Sampling: Surveying Political Leaders
Gregory G. Holyk, Washington and Lee University (holykg@wlu.edu)

This paper examines diversity in sampling techniques, specifically, the usefulness of non-probability sampling to survey the attitudes of political leaders, a specialized population that is notoriously difficult to access. Due to the small population and difficulties with access and cooperation it is not possible to select a random sample of political leaders. Political leaders have enormous power and influence in the political process and the study of their characteristics, methods of decision-making, behaviors, and attitudes and belief systems are essential to understanding the political process. How can we get a somewhat representative estimate of the attitudes of political elites? I examine the strengths and weaknesses of the Chicago Council on Global Affairs' "Global Views" leader surveys, which they conducted from 1974 to 2004. The surveys employed quota sampling of leaders in different areas using comprehensive lists of membership in each type of position. Although inferences based on these samples are tenuous at best, I argue that this approach to surveying political leaders results in a reasonable estimation of leader attitudes that is preferable in comparison to other available methods -- analysis of political writings and actions of political leaders (e.g., Barber, 1992; George, 1980), semi-structured interviews of individual leaders (e.g., Kingdon, 2002), direct observation of the behaviors of select leaders (e.g., Fenno, 1977), analysis of the statements of leaders in public statements, newspapers, or other media (e.g., Ragsdale, 1984; Canes-Wrone, 2001), or examination of actual policies and vote -- when the goal is a comparison of leader and public attitudes.

Collection Research for Household Surveys at Statistics Canada
Helene Berard, Statistics Canada (helene.berard@statcan.gc.ca); Milana Karaganis, Statistics Canada (milana.karaganis@statcan.gc.ca); Francois Laflamme, Statistics Canada (francois.laflamme@statcan.gc.ca)

One of the continuing challenges for a survey organization such as Statistics Canada is to explore ways to carry out efficient survey designs that take advantage of new collection technologies; that satisfy cost, quality and operational constraints; and that reduce respondent burden. A variety of collection research projects and new initiatives are carried out at Statistics Canada to help meet this challenge.

This paper starts with a brief description of the household survey framework at Statistics Canada. It is followed by a review of the types of collection research projects and new initiatives that are planned and carried out for household surveys at Statistics Canada in order to reduce non-sampling error (coverage error, nonresponse error, and measurement error) and improve collection operations efficiency (for example through the use of responsive designs). It concludes with a discussion of future challenges and directions for our collection research program.

Dichotomous Voting Intention Questions and Response Order Effects: Individual Participant Data Meta-Analysis.
Alexandre Morin Chassé, Université Laval (alexandre.morin-chasse.1@hotmail.com)

Are dichotomous voting intention questions subject to response order effects? To arrive to a clear conclusion, I conduct a quantitative systematic review of available datasets on the 2008 election, develop the tug-of-war framework to help analyzing results, conduct meta-analyses of aggregated data, review the hypotheses in the literature, develop a statistical shortcut to finally perform data mining of individual participant data using multivariate additive and multiplicative regressions. No hypotheses are solidly confirmed and exploration reveals nothing consistent.

At this point, it is still possible that a theory that escaped my review or some developments to come may unearth a subgroup more likely to be subject to response order effects on such questions. A last experiment narrows this gap to a small and unlikely eventuality by providing a positive direct test of the null hypothesis. This research leads us to a strong negative conclusion : there are no response order effects on dichotomous voting intention questions for presidential elections in surveys conducted via phone interviews.

Significant Factors Governing the Use of Auditory Stimuli in Web Questionnaires
Sentagi S Utami, University of Michigan (sentagi@umich.edu); Rahmad Dawood, University of Michigan (rahmadd@umich.edu); Mojtaba Navvab, University of Michigan (moji@umich.edu)

There has been a trend in the use of auditory stimuli in web survey. However, its use is still considered as novel method and available references are limited. This paper identifies six important factors that can affect survey error rate and need to be anticipated when using auditory stimuli in web questionnaires. These factors were derived from a pilot survey as part of a room acoustics study where only 80 out of the 160 invited subjects completed the survey.

The six factors identified are: (1) computer hardware, (2) operating system, (3) Internet connectivity, (4) how familiar the subjects are with the survey content and terminology, (5) complexity of the survey interface, and (6) space where the survey was completed.

The first factor is related to the type of computer and listening device that the subject is using. Since different operating system and software environment handles and processes auditory stimuli differently, to generate the stimuli may require additional software. Also since the auditory stimuli are streamed, the speed and reliability of the subject's Internet connectivity is a significant factor to produce an uninterrupted audio flow. The characteristic of the used passage for the auditory materials more significantly affected the assessment of the speech intelligibility than the need to have trained subjects for the particular study. Since the objective of using auditory stimuli is for hearing assessment, background noise where the survey is being completed can have significant affect on the final listening experience.

These findings were then used to create a new web questionnaire; where a usability study was then conducted to evaluate how effective this new design is in resolving the above issues.

The Financial Crisis and the Bailout: Frame Analysis of the Chicago Press.
Fabrizio Ceglia, University of Illinois at Chicago (fabrizio.ceglia@gmail.com); Bruno Carriço Reis, Universidad Rey Juan Carlos (reysbr@yahoo.com.br)

In this study, we analyze one of the most debated issues of this financial crisis: the financial bailout. In fact, this represents the perfect case of study, focusing the attention of public opinion and influencing the choices of politicians in the middle of the campaign.

By a news framing analysis of the Chicago press, this research studies the valuation that journalists made of the bailout. In particular we compare the coverage of the two most important newspapers of the State of Illinois: The Chicago Tribune and the Chicago Sun-Times. The choice of the Chicago press was due to the important role played by the city of Chicago in 2008 presidential election. The presence of Obama, junior United State Senator from Illinois and candidate for president of US, increased the attention of the Chicago media for the political and economic issues related to the electoral campaign.

Another approach we use to explain how officials could easily impose their frames to the media, using their powerful position as privileged sources and interpreters of the facts, is the model of cascade activation of Entman. Even if Entman used his model to investigate the coverage of foreign policy, we think that could be a good model also for this kind of analysis. In fact, as in the cases studied by this scholar, we analyze a context of crisis where media is generally more dependent on the official sources. Finally, we try to determine whether the interpretation of the media on a particular issue gives priority to the audience opinion or, on the contrary, gives more credit to the position of the political leaders. For this reason, we compare the analysis of the press with surveys realized in the same period of time, describing the aggregated public opinion.

Framing effects and racial prejudice: The moderating role of need for cognition.

Porismita Borah, University of Wisconsin-Madison (borah@wisc.edu); D. Jasun Carr, University of Wisconsin-Madison (dj carr@wisc.edu)

Research on framing has begun to expand beyond the simple examination of framing effects and has started delving into those conditions that might amplify or abate them (Druckman, 2001; Druckman & Nelson, 2003). Recent studies have examined what impact moderators, such as political knowledge, need for cognition, and need to evaluate, have on this process (Druckman & Nelson, 2003, de Vreese, 2004; Schuck & de Vreese, 2006).

The present study builds on this trend through the use of an online experiment examining the role of need for cognition, generally defined as a tendency to “engage in and enjoy thinking”, as a moderator (Cacioppo & Petty, 1982). The online experiment dealt with the issue of civil liberties conflict and consisted of both pre and post-manipulation survey items. Following the pre-manipulation survey, respondents were presented with experimental manipulations of a news story constructed using either a public safety or a free speech frame. Following this exposure, respondents’ attitudes towards a KKK rally to be held in the University campus, and racial prejudice, were assessed through a number of survey items (including the modern racism scale to measure the resulting racial prejudice).

Results show that participants’ attitudes toward a possible KKK rally to be held were affected by the frame used irrespective of participants need for cognition ($p < .001$), in line with prior studies (Nelson et al., 1997). However, need for cognition moderated the effects of the frame on racial prejudice ($p < .05$), such that individuals with high levels of need for cognition and who were exposed to the public safety frame showed lower levels of racial prejudice. The moderating role of need for cognition demonstrated in this study have significant implications for psychological processes involved in framing effects.

Fill in the Blank: SED Verbatim Response Specificity as a Function of Perceived Audience.

Lino M Jimenez, NORC (jimenez-lino@norc.org); Lindsay Virost, NORC (virost-lindsay@norc.org); Vincent Welch, NORC (welch-vince@norc.org)

Survey researchers strive to obtain the most accurate information from respondents. One method to obtain accurate information is to provide respondents the opportunity to answer survey items as open-

ended, verbatim responses. A problem with verbatim responses is that they allow respondents varying degrees of specificity when answering, which can change the accuracy of the information.

Another problem affecting the accuracy of survey information is the survey respondent's intended audience. Survey participants have in mind an audience when responding to question items. This perceived audience can change how participants respond.

In the current study, we examine the results of a quasi-experiment comparing different groups' responses to verbatim items. We use data from the Survey of Earned Doctorates (SED). The SED collects data on multiple characteristics of new doctorate recipients and is approximately 40 items in length. We collected and coded verbatim data for 3 survey rounds of a verbatim item that asks participants to list the organization and geographic location of where they will work or study.

We anticipate verbatim responses of international respondents who have foreign, future work-plans to be less specific than domestic respondents as a function of perceived audience. Also, we expect domestic respondents who have future plans in different states to have similar responses as other domestic respondents within state. Mode-effects (web, PAPI, CATI) will also be examined.

Searching for Alternatives: Findings from Recent Mail Surveys by the Asahi Shimbun.

Nicolaos E. Synodinos, University of Hawaii (nick@hawaii.edu); Eiji Matsuda, The Asahi Shimbun (matsuda-e@asahi.com)

Surveys sponsored by the Japanese news media were predominantly face-to-face until the early 1990s. Thereafter, they were replaced with directory-based telephone interviews and since the early years of the millennium have been mostly RDD-based. As elsewhere, response rates (of all modes) decreased and are a major concern.

Besides telephone and personal interviews, the Asahi Newspaper has been investigating alternative approaches (e.g. mail and web). This paper concentrates on findings from three recent Asahi nationwide mail surveys that differed in topic (trust; health; and politics) and timing (early 2008; mid 2008; and early 2009). Each survey drew a random sample of 3,000 from the Voter Registration Lists and all three questionnaires consisted of 7 pages of questions. Their implementation procedures were similar consisting of a notification postcard mailed one week before the questionnaire, a reminder postcard mailed 2 weeks later to non-respondents, and followed with a replacement questionnaire 8 days later. The original mailing had an included incentive consisting of a ballpoint pen (retailing for about 500 yen). In addition, respondents were promised a book coupon of 1,000 yen upon returning the questionnaire.

Despite the different topics, the response rates (RR2) of these three surveys were similar and remarkably high: 78% (trust), 77% (health), and 79% (politics). These high rates can be attributed to utilizing (with some modifications) various principles of Dillman's Tailored Design Method and the well-known sponsoring organization.

These and other Asahi mail surveys found no differences in response rates by educational level or degree of urbanization. As other methods, mail surveys tend to under-represent persons in their 20s; but, they are substantially better than RDD and equal or better than face-to-face interviews.

The results are extremely encouraging for those allowed access to official lists for sampling purposes. Other researchers should be encouraged to explore available ABS frames.

The Use of Technology in Seat Belt Survey Data Collection.

Larry N. Campbell, RTI International (campbell@rti.org)

It is always tempting to believe newer technologies are better than the old ones, where we attempt to mold them into solutions to old problems. This paper offers a comprehensive review of the technologies currently used for data collection in federally mandated state-level seat belt surveys, and evaluates the diversity of new technologies used in other data collection efforts that could be employed in these and other highway safety surveys. We review and compare these new technologies to the old technologies to determine if gains in data quality, efficiency, and cost savings would be achieved by making the technology upgrade.

Currently, most state seat belt surveys use paper and pencil as their mode of data collection, where data is entered using scannable forms or manual data entry. Voice recording technology is used in some rural states with low traffic volume, such as Alaska. This paper explores the possibility of using new technologies for data collection, such as PDAs and GPS, to improve data quality, reduce data loss, reduce costs, and improve efficiency. For instance, data collection sites can be rural, or may lack clear road identifiers. We evaluate using GPS embedded photographs to aid data collectors and ensure data is collected at the correct site, and we also evaluate its cost implications.

An important consideration when discussing new technology in data collection is the type of data collectors employed. Would technology be welcome among the field observers? In some states the field observers are predominantly senior citizens. How might data collection be affected in such states? We discuss training and support for field staff with newer technologies. Are there other challenges for this population and current data collection methods? This research answers these questions.

Item Nonresponse Analysis for a Mixed-Mode Survey.

Boris Lorenc, Statistics Sweden (boris.lorenc@scb.se); Kristoffer Olsson, Stockholm University (kristoffer.80@spray.se)

The literature is equivocal on the existence of a systematic difference in item completion rates between the mail (postal) mode and the web mode, two self-administered data collection modes. The study addresses that issue in the context of a concurrent mixed-mode contact strategy with self-selection of the data collection mode. However, the study should be viewed as exploratory rather than confirmatory. In order to balance the differences in demographic properties of the respondents, who were free to choose the mode for their participation, a propensity score weighting was used; and in order to take into account the differing properties of the items comprising the questionnaire, a logistic regression was applied. The results show a small but significant difference in base item completion rate in favour of the web mode, not attributable to any of the investigated item properties, demographic properties or data collection procedures. Additionally, several item level factors were found to differentially influence item completion rates in the two modes, indicating an interaction between item and mode—some of them with direct applicability to data collection improvement while others awaiting further work in order to understand them better.

The Effect of Removing Cell-phone-only (CPO) Weighting

Yelena Pens, Arbitron (yelena.pens@arbitron.com); Kelly Dixon, Arbitron (kelly.dixon@arbitron.com)

Arbitron Inc. has developed the Portable People Meter (PPM™), a new technology for media and marketing research. A panel-based methodology is used to collect PPM™ related data. Arbitron is

committed to increasing the percent of CPO panelists in PPM markets. Bringing the CPO panelist percentage closer to the population percentage reduces the need for CPO weighting, since the group will not be under-represented.

Arbitron has been using CPO/landline status as a weighting variable in all PPM markets. Arbitron has been improving methods to sample CPO panelists, and studying the effects of CPO panelists on the estimates. However, Arbitron was concerned about the quality of the population estimates, and how variation in population estimates is affecting the weightings of CPO panelists. The concern for improving the reliability of the estimates for CPO panelists has been on-going.

In 2009, Arbitron conducted an analysis to measure the ratings impact of removing CPO weighting in PPM markets. In the analysis, Arbitron compared estimates that did not include CPO weighting to the current methodology of estimates that included CPO/Landline weighting.

In this paper, Arbitron presents the results of the ratings impact based on the proposed methodology to the current methodology. The estimates were calculated by station and format level for various demographic groups. Additionally, an analysis was conducted looking at all the weighting variables to determine listening variation. Recommendations and future research will also be presented.

The Discriminating Voter: An Analysis of the Demography and Geography of Discrimination

Paul G Harwood, Public Opinion Research Laboratory, University of North Florida

(pharwood@unf.edu); Nicholas J Seaton, Public Opinion Research Laboratory, University of North Florida (nicholas.seaton@unf.edu)

Although America elected its first African-American president, the passage of propositions from California to Florida simultaneously and profoundly limited the rights of gay Americans, illustrating that the “blessings of Liberty” expressed in the Constitution are yet to be attained by gay America, in spite of the noticeable strides made by other historically marginalized minority populations (i.e., women, African-Americans) ... particularly in this most recent election cycle.

Utilizing a nationwide survey (N=1089, conducted October 2008), we examine quantitative data to analyze the demographics of discrimination by comparing three different survey questions gauging which demographic factors present in a political candidate would most influence a respondent's vote choice. When asked whether race, gender, or sexual orientation would impact their choice of candidate, twice as many respondents reported that sexual orientation would impact their vote choice than the gender and race of the candidate combined. We distill these figures and isolate the demographic factors present among the respondents that most impact these outcomes by conducting a series of logistic regressions and probabilities on each of the three questions in the survey.

Additionally, we break down the analysis by geography to determine whether regional differences exist, particularly in areas that have codified either legal provisions or prohibitions of gay marriage and/or adoption. By parsing the data by region and/or by state, we are able to isolate geographic trends to determine whether legislation or court decisions made within the area as regards the rights of sexual minorities impact public opinion.

While assessing public opinion on attitudes toward race, gender, and homosexuality are certainly not novel concepts, in this paper, we examine not only their salience in this most recent presidential election in relation to other polemic social issues, but also by isolating prejudice based upon a series of demographic variables.

A Multi-Mode Approach for Leave-Behind Data Collection of Actigraphy in a Population Survey
Jane L McPhillips, NORC at the University of Chicago (mcphillips-jane@norc.org); Katie Lundeen, NORC (lundeen-katie@norc.org); Angela Jaszczak, NORC (Jaszczak-Angie@norc.org); Stephen Smith, NORC (smith-stephen@norc.org)

Additional data collection materials left behind with respondents after the completion of the main survey, such as questionnaires, diaries, and the collection of biospecimens have long been a part of survey research. Respondents are increasingly being asked to participate in 'leave-behind' measures and survey researchers must carefully track their response rates and data quality. NORC recently conducted a pretest of the Second Wave of the National Social Life, Health and Aging Project (NSHAP), an innovative, multi-modal study of older adults born between 1920 and 1947 that examines the interaction between aging, social relationships, and health outcomes. The NSHAP interview consists of an in-person computer assisted interview and a series of respondent-administered leave-behind measures, including actigraphy.

Actigraphy is a minimally-invasive method used to measure sleep and activity cycles. To obtain data, respondents were asked to wear a wrist-watch device and complete a sleep diary for 72 hours to obtain a continuous recording of sleep and activity over three days. Although prior surveys (CARDIA, NHANES) have collected actigraphy measures, a gap in the literature exists for methods to achieve high response rates and data quality when integrating leave-behind measures with in-home data collection.

In the NSHAP pretest, 109 out of 120 respondents agreed to participate in the actigraphy measure. This paper describes our approach to ensuring high response rates and data quality for the leave-behind measure. We employed a multi-modal approach that included in-person, mail, telephone prompting, and a self-administered questionnaire. We will provide information on our return rates, number of prompts, average days to complete the measure, timely or delayed returns of the materials, respondent characteristics and overall data quality methods. We will discuss lessons learned for implementing the measure in the main study, and more generally, for other studies exploring the use of more extensive post-survey measures.

Data Collection from Third Parties: The Rental Agents Survey Case Study
Marilyn Worthy, Energy Information Administration (Marilyn.Worthy@eia.doe.gov)

The Energy Information Administration (EIA) conducts an in-person household survey called the Residential Energy Consumption Survey (RECS), to collect energy data on U.S. households. Householders who own their homes are able to report energy characteristics about their home since they are responsible for the operation and maintenance of their energy systems and pay their own utility bills. They observe direct relationships between type of equipment, fuel, maintenance, and energy expenses. However, about 30 percent of U.S. households rent their homes. Householders who rent their homes are not able to observe these relationships because landlords are responsible for energy system maintenance, and utility costs are often included in the rent. Thus EIA supplements RECS by conducting a concurrent Rental Agents Survey (RAS) to ask rental agents and landlords about energy-related characteristics for home renters who do not pay their own utility bills. The RAS is much shorter than the RECS and focuses on information which renters often cannot provide, such as the type of equipment used to heat the home or the fuel for water heating. This paper will describe how data from rental agents and landlords is collected and used to improve EIA's estimates for rented units. This application also offers an alternative survey methodology to other organizations that collect survey data where third-party data may be more accurate than data obtained from the initial respondent.

The Influence of Cultural Dimensions on Survey Response Styles: An Analysis of IBM Global Employee Surveys for India.

Randal Robert Ries, IBM (rries@us.ibm.com)

As one of the world's largest globally integrated enterprises IBM regularly conducts global employee opinion research on topics ranging from job satisfaction and IT delivery to understanding of the company's values. Analysis of these global survey results at the country level provides critical insights for various corporate stakeholders including corporate communications programs, IT delivery, and human resources. The results of IBM's internal research is consistent with accepted understandings of cultural influences and behaviors in survey taking for the U.S., Japan, Germany, and other developed markets. For example, it is generally accepted that in Japan modesty and "politeness" influence survey responses and cause a tendency toward the middle of the scale being used. A similar understanding survey response styles for India—where IBM's employee population has increased over 400% since 2004—does not currently exist.

IBM's employee satisfaction scores for India on a variety of measures on various global surveys skew disproportionately positive in comparison to other IBM countries. For example, when asked to rate their overall satisfaction with IBM's corporate intranet, nearly half (48%) of employees in India responded "very satisfied" or at the top of the scale. This top box satisfaction level is nearly double that seen in the U.S. (27%), seven times that seen in Germany (7%), and twelve times that in Japan (4%). These results and others for India on IBM internal surveys appear to be indicative of Extreme Response Style (ERS). A definitive and comprehensive study on India and survey response style has not been undertaken by either academic or professional market researchers. This paper will apply published research on dimensions of culture, cross-cultural relations, and large global organizations to an in-depth analysis of IBM's global survey results in order to provide insights useful to further research conducted in the Indian market.

Imputation of Income, Poverty, and Medicaid Status in a Survey of Health Insurance and Access to Care.

Tom Duffy, ICF Macro (thomas.duffy@macrointernational.com); Ronaldo Iachan, ICF Macro (ronaldo.iachan@macrointernational.com); Sara Bausch, ICF Macro (sara.bausch@macrointernational.com); Bo Lu, The Ohio State University (blu@cph.osu.edu); Tim Sahr, Health Policy Institute of Ohio (trsahr@hpio.net)

The Ohio Family Health Survey (FHS) is a telephone survey of the health and health insurance status of adults and children in Ohio. The FHS prescribed confidence intervals for estimates of insurance status for several population subgroups: rural regions, ethnic minorities, families in poverty, etc. This paper presents and assesses the imputation methods developed for the survey with special attention to the imputation of income and Medicaid status, two inter-related variables with a substantial amount of missing data. Both Medicaid status and income were imputed using multivariate regression models for ordinal and binary outcome variables (Proc MI). The predictors included the following variables: Gender, Age, Education, Race/ethnicity, Tenure (Own vs. Rent), Employment status (full time versus not), Insurance status, Household size, and Marital status. We developed 5 models.

Coding Complex Responses.

Linda G Kimmel, Michigan State University (kimmell2@msu.edu); Jon D Miller, Michigan State University (jdmiller@msu.edu)

Each year, the Longitudinal Study of American Youth (LSAY) includes a number of open-ended items in its yearly survey. The coding procedures used vary depending on the nature of the question. Broadly, we use three different types of open-ended questions.

The easiest responses to code are short answer responses such as the names of magazines or television shows. These responses are coded with the assistance of Microsoft Access look-up tables. We will give examples of one such look-up table, as well as the procedures used to update the table.

Second, are open-ended knowledge questions such as the meaning of DNA and molecule. The responses are often lengthy, and require complex judgments by coders. We will describe in detail the training procedures used, as well as the methods used to assess inter-coder reliability.

The final type of open-ended response is the participants' occupation. These responses would seem to lend themselves to an automated coding process. However, simple job titles such as nurse or engineer can be used for a variety of different occupations. Occupation is coded based on the response to the occupation question in conjunction with a variety of other information including highest level of education and college major. We will provide examples of how multiple sources of information are used to clarify responses.

Open-ended questions provide the LSAY with valuable information about a wide range of topics including the participants' careers, their use of various media, and their scientific knowledge. While coding of the responses can be labor-intensive, we find the results to be worthwhile, and hope that our experiences will prove useful to others who are looking for procedures to code open-ended questions.

Variance of the 2010 Census Coverage Measurement Logistic Regression based Dual System Estimator of Population Size.

Richard Griffin, U.S. Census Bureau (richard.a.griffin@census.gov)

Survey estimates that are functions of logistic regression parameters are much more common now than in the past. However, such estimators require a different approach when planning for the accuracy of the estimates for a proposed sample design.

For the 2010 Census Coverage Measurement (CCM) Survey, the dual system estimates (DSE) will be a function of estimated logistic regression parameters. For previous coverage measurement surveys, which used traditional DSE (defined explicitly as a function of other statistics from the sample), Taylor Linearization methods were used to approximate the variance of DSE for sample design planning purposes. However, for 2010 CCM, our estimator is not defined explicitly as a function of other statistics from the sample. Taylor Linearization variance estimation techniques applied to an estimate that is a function of sample statistics is not feasible.

Binder (1983) presents Taylor Linearization methodology applicable when the estimator cannot be defined explicitly as a function of other statistics from the sample. He shows how these results can be applied to logistic regression

This paper describes the general theory for logistic regression from Binder (1983). A detailed example for a simple logistic regression model with only one independent variable plus an intercept term is provided. The Census Bureau investigated the increase in variance if the planned sample size were to be cut. The likely reduction was such that selected racial minority groups would have no cut while other groups were cut at a single rate less than or equal to 50%. The theory of this paper is applied to this case and the

resulting increase in variance from the sample size reduction is compared to empirical results using jackknife variance estimation.

Measuring the Effects of a Nationwide Community-Based Participatory Design on Outcomes Related to Electronic Health Information Exchange.

Stephanie Rizk, RTI International (srizk@rti.org); Robert Bailey, RTI International (rbailey@rti.org)

The purpose of this research is to determine the impact of a community-based participatory design on developing solutions to various challenges posed by privacy and security issues related to electronic health information exchange (HIE). Between 2005 and 2009, the U.S. Department of Health and Human Services (HHS) funded the Privacy and Security Solutions for Interoperable Health Information Exchange project. The purpose was to better understand variation in business practices, policies, and laws related to privacy and security and develop effective solutions to interoperable HIE. Developing solutions to complex privacy and security related issues is frequently cited as a serious challenge to HIE, and differing opinions about how to implement solutions threaten to impede progress. Rather than collecting traditional data from respondents, the project employed a large-scale, nationwide community-based participatory design that sought to engage stakeholders in implementing consensus-based solutions.

To establish the scope and extent of the challenges, we will draw on baseline business practice data that were submitted by each state using common data collection tools. We will then review the number of participants engaged in the effort as reported by each state (n=34) to determine the extent of community participation. Finally, we will review the number and scope of various outcomes (guidelines, templates, executive orders, legislation) to draw conclusions about the effect of this unique methodology on successfully accomplishing the goals established at the outset of the project.

By July 2009, the project supported development of over 150 documents and tools which have been made publicly available to entities seeking to develop their own solutions to privacy and security related challenges. This supports the hypothesis that the design led to significant outcomes, and achieved added benefits, such as establishing a dialogue between entities that will support successful implementation of many of the HITECH initiatives.

Optimizing Call Time in an RDD Survey: BRFSS Experience

Sean Hu, Centers for Disease Control and Prevention (shu@cdc.gov); Brian F Heas, RTI (bhead@rti.org); Lina Balluz, Centers for Disease Control and Prevention (lballuz@cdc.gov)

Over the past three decades response rates have declined significantly, and simultaneously costs associated with running a telephone survey have increased. Survey researchers have attempted to mitigate the decline in response rates and reduce the costs to complete cases with advance letters, prepaid incentives, and scripted answering machine messages. These methods have met mixed results. Much less attention, however, has been given to the impact of call scheduling, and in particular the degree to which there exist temporal and seasonal variations in optimal call scheduling. We use detailed call history data from 2008 Behavioral Risk Factor Surveillance System (BRFSS), the world's largest on-going public health telephone survey, to assess the efficiency of our calling protocols. The purpose of this research is to identify ways to reduce interviewer effort by calling more efficiently, and determine whether there is a relationship between BRFSS calling schedules and BRFSS outcomes rates. Attempt data from eleven BRFSS states are used to test the hypothesis that there is an interaction between time-of-year and time-of-day of first attempt that affects the number of attempts to reach a final disposition code. The hypothesis that a seasonality effect is dependent on time-of-day is tested using AAPOR outcome rates

(CON1, COOP2, RR2, and REF1) from detailed call history data on first attempts from a subsample of eleven states.

AAPOR Concurrent Session C

Saturday, May 15, 2010

2:15 p.m. - 3:45 p.m.

Attitudinal Data Collected via Cell Phone.

"Cell Mostly" or "Cell Mainly": What is More Representative?

Brian Shepherd, NORC (shepherd-brian@norc.org); Megha Shah, NORC (shah-megha@norc.org); Wei Zeng, NORC (zeng-wei@norc.org); Karen Wooten, Centers for Disease Control and Prevention (kgw1@cdc.gov)

As dual frame random-digit dial landline and cell telephone surveys proliferate, research into best methodological practices becomes increasingly necessary. A common task for many cell telephone surveys is screening for households without a landline or that rarely use their landline. Interviewing individuals in landline-only/mostly households is critical to improving landline RDD coverage. The question of who is "mostly" a cell telephone household is a difficult one. Ideally, we wish to capture all households that would be excluded from an RDD survey, despite having a landline. Researchers have typically worded questions to identify these households in one of two ways: One form of this question has focused on the likelihood of answering a call placed to a landline while the other form of this question has focused on the relative usage of the household's landline versus cell telephone. In a recent pilot study of cell telephone users on the National Immunization Survey (NIS) both questions were administered to each respondent. In this presentation, we examine responses to these two questions. We seek to answer a number of questions: How often do the answers to these questions suggest different "cell mostly" classification? When responses do vary, does the data suggest that one question is more effective than the other? Do different demographic groups tend to respond to these two questions differently? We examine each of these questions as well as compare key demographic and health indicators to both the NIS landline survey and the National Health Interview Survey (NHIS).

Measuring State Health Trends--Impact of the Cell-only Population

Frederica Conrey, ICF Macro (frederica.conrey@macrointernational.com); Randal ZuWallack, ICF Macro (randal.zuwallack@macrointernational.com); Kristie Hannah, ICF Macro (kristie.hannah@macrointernational.com)

In order to evaluate the impact of the Centers for Disease Control's (CDC) recommendations and funding programs for HIV prevention, the CDC is conducting the Assessment of HIV Testing in Clinical Settings (AHITS). Prevalence estimates for HIV testing are available from the National Health Interview Survey (NHIS), but only the Behavioral Risk Factor Surveillance Survey (BRFSS) offers state-level estimates.

In 2003, the percentage of adults tested for HIV was 45.9% according to the BRFSS, while the NHIS measured 40.2%. By 2008, the NHIS rate had climbed to 44.6% while the BRFSS rate declined to 39.9%. During this time of divergence, the percentage of adults living in cell-only households increased dramatically. According to a 2008 cell phone survey, 49% of cell-only respondents have ever been tested for HIV. As a telephone survey, the BRFSS is susceptible to bias as a result of the exclusion of cell-only households.

In 2009, the BRFSS added a cell phone sample, but in the interim, wireless substitution has disrupted the trend in HIV testing. To adjust for the cell-only bias, we developed a national propensity score model that measures the probability of observing an adult in BRFSS versus NHIS conditional on having been tested for HIV. We apply the model to the BRFSS to correct for the decrease in HIV testing rates attributable to the exclusion of cell phone only adults.

Implications for correcting other national surveillance instruments for the purpose of trending are discussed.

Cell Phone Usage and Differences in Behaviors Surrounding the H1N1 Outbreak

Eran N Ben-Porath, SSRS/Social Science Research Solutions (ebenporath@ssrs.com); Gillian K Steelfisher, Harvard School of Public Health (gsteelfisher@gmail.com); Melissa J Herrmann, SSRS-Social Science Research Solutions (mherrmann@ssrs.com)

This paper studies the differences in behaviors and attitudes between cell phone only (or mostly) users and people answering land line phones – surrounding the 2009 outbreak of H1N1 influenza. Existing research on the unique characteristics of cell phone only users indicates they are more likely to engage in riskier health behaviors, such as excessive drinking, and less likely to engage in proactive health behaviors such as immunization. These behaviors are believed to extend beyond the known demographic differences between cell phone only users and landline users, namely, age. Based on a series of ongoing surveys, the present study tests the hypothesis that cell phone only users will be less likely to engage in behaviors that are designed to minimize the risk of contracting H1N1 influenza ('swine flu'). The findings are consistent with this expectation on various measures: cell phone only users were less likely to report behaviors such as using hand sanitizers, reducing physical contact with others or avoiding areas such as large shopping areas. Controlling for age, does not eliminate these differences. In fact, the differences are most pronounced for older cell phone users (age greater than 29) than for the youngest age group. Two exceptions to this stand out: at the early stages of the outbreak, the differences in behaviors were strongly associated with Latino heritage and geographic location, and no association with phone usage was observed. In addition, the intention to get the H1N1 vaccine did not seem to vary across phone usage categories. Overall, the findings we report indicate that the inclusion of cell phone only respondents in public health surveys is necessary to assure that the divergent health behaviors associated with the phone usage categories are represented in the survey sample.

Comparison of Response Measures Between Landline and Cell Phone Modes in Behavioral Risk Factor Surveillance System

Mohamed G Qayad, Centers for Disease Control and Prevention (MQayad@cdc.gov); Pranesh Chowdhury, CDC (pchowdhury@cdc.gov); Machell Town, CDC (mpt2@cdc.gov); Lina Balluz, CDC (Lballuz@cdc.gov)

Behavior Risk Factor Surveillance System (BRFSS) is a state-based landline telephone survey conducted in all U.S. states/territories throughout the year. Due to increase in cell phone usage, 48 states collected data from cell phone only households in 2009. We used landline and cell phone only household data from 2009 BRFSS and compared the following response measures: Council of American Survey and Research Organizations (CASRO) response and cooperation rates, percent completed the screening questions, and interview completion and refusal/termination percent among responders of screening questions. The median CASRO response rate for the landline was 52% and 39.5% for cell phone only households. The median cooperation rate for the landline was 74.7% as compared to 77.4% for cell phone only households. The median percent completed the screening questions was 15.9% for the landline and 5.5% for the cell phone only households. The median percent completed the interview was

66.5% for the landline and 77% for the cell phone only households. The median percent who refused or terminated the interview was 20% for the landline and 15% for cell phone only households. The cell phone only households had a lower percent of people who refused or terminated than the landline. Also, cell phone only households are less likely to complete the screening questions but are more likely to complete the interview once they complete the screening questions.

Risky Business: A Comparison of Traffic Safety Behaviors by Telephone Status in a National Dual Frame Design.

John M. Boyle, Abt SRBI Inc. (j.boyle@srbi.com); Faith Lewis, Abt SRBI Inc. (Lewis@srbi.com); Brian Tefft, The AAA Foundation for Traffic Safety (BTefft@AAAFoundation.org)

Cell phones have produced a dramatic decline in household coverage in telephone surveys using RDD landline sample. According to recent National Center for Health Statistics (NCHS) estimates, more than one in every five American homes (20.2%) had only wireless telephones during the second half of 2008. Additionally, 24.4% of households with both wireless and landline telephones received all or almost all of their calls on wireless telephones.

Aside from demographic differences between traditional landline RDD samples and cell phone only samples, NCHS has found differences in risky health behaviors, such as binge drinking and smoking. These differences in risky behaviors persist after weighting for demographic characteristics. This paper investigates whether these differences between cell and landline samples extend to other forms of risky behavior.

The 2009 Traffic Safety Culture Index Survey provided an opportunity to test differences in driving safety behaviors in a dual-frame national sample of households. The first frame consisted of a total of 1500 completed telephone interviews with adults aged 16 and older from a national random digit dialing (RDD) sample of households with landlines. The second frame consisted of 1001 completed interviews with adults aged 16 and older from a national sample of cell phone banks.

This paper compares reported rates of speeding, texting, driving without a seatbelt, driving after drinking, running red lights, motor vehicle accidents and tickets for moving violations between cell and landline samples. These rates are compared between cell only, cell mostly and other landline respondents from the dual frame survey. The findings suggest that traditional landline RDD surveys will underestimate driving risk behaviors, as they have been found to do with alcohol and tobacco use. Hence, we find it is risky business not to use a dual frame design in studies of risk behaviors in the American public.

Election Attitudes & Polls

Do Voters Live Vicariously Through Election Results?

Adam Sage, RTI International (asage@rti.org); Bonnie E. Shook-Sa, RTI International (bshooksa@rti.org)

To date, it is unknown whether voters react to elections as if the outcome is a personal victory or defeat. Affect Control Theory (ACT) suggests that people strive to maintain consistency between expectations and experiences by responding to events emotionally. INTERACT is software that embodies ACT's mathematical formulas for calculating quantitative values for the discrepancies between expectations and experiences, and allows for a mathematical simulation and interpretation of events.

This paper uses data from the American National Election Study Time Series (ANES) to examine how discrepancies between pre-election expectations and the outcome of the 2008 presidential election predict voters' post-election intentions of engaging in political action (e.g., attending town hall meetings, signing petitions, joining a protest, etc.). The ANES is a nationally representative study on voting behavior and public opinion that occurs every 4 years concurrent with each presidential election. This study is conducted in two waves during which respondents are interviewed before and after the election, with no knowledge that they will be contacted for a post-election interview. For the 2008 ANES, conducted by RTI International, respondents were asked questions pertaining to their outlook on the election and preferences for the presidential candidates in the pre-election survey and about their intentions to engage in political action in the post-election survey.

Using respondents' predictions of the election outcome and their voting intentions captured during the 2008 ANES and the actual 2008 presidential election outcome, we create a discrepancy variable via INTERACT. We then use multivariate regression analyses to examine the relationship between the discrepancy variable and measures of intended political action, including making political donations, signing petitions, and distributing propaganda. This research will further our understanding of how voters internalize political processes and how political events influence involvement in political issues.

The Development of Conservatives and Liberals: An Analysis of Political Socialization and Change.

Jon D. Miller, Michigan State University (jdmiller@msu.edu); Jason Kalmbach, Michigan State University (kalmba11@msu.edu); Linda G Kimmel, Michigan State University (kimmell2@msu.edu)

The American political system is becoming increasingly polarized by ideology. It is important to understand the factors that are producing this change. The Longitudinal Study of American Youth (LSAY) has been following the same cohort of students since 1987. The participants are now in their mid-30's and more than 70% continue to participate in the LSAY (N = 3700). This paper will use the extensive longitudinal record for each young adult (including telephone interviews with their parents during their high school years) to examine the transmission of parental values and political perspectives to the new generation. The longitudinal record on education and employment will be used to assess the impact of educational and occupational socialization factors to the development of current political values and attitudes. Geographic factors will also be included in the analysis (red states, blue states, battleground states). The paper will utilize a set of structural equation models to estimate the impact of each of several factors on the development of ideological partisanship by LSAY young adults and the influence of their ideological partisanship on their vote in the 2008 presidential election. The paper and presentation will include sufficient descriptive materials to attendees who are not conversant with these model building techniques. Philip Converse is advising on this analysis, but will not be a co-author of the proposed paper.

2008 Election Exit Poll Design Effect

Clint W. Stevenson, Edison Research (cstevenson@edisonresearch.com)

For the 2008 election exit poll it is impossible to select a simple random sample. Consequently, a cluster sample is required in this instance. In surveys that employ complex sample designs the standard errors are larger than a simple random sample. Furthermore, standard errors from individual questions from the questionnaire may be larger depending on the clustering effect for that question. This paper will present the standard error from the exit poll questions asked during the 2008 election exit poll assuming a simple random sample. The standard error will then be recalculated for these questions by applying the design effect. Presented here will be the final design effect for the 2008 election exit poll for each of the states

using questions from the questionnaire. To complement these results the design effects are compared to previous exit polls from the 2004 general election. The process for calculating the design effect for the election exit poll incorporates the delete-1 jackknife. This paper presents a procedural discussion of the calculation of the design effect and also presents an alternative approach on how the design effect can be calculated for exit polls.

Electing the President 2008. A Longitudinal Web Panel Survey from November 2007 to November 2008.

Nat Ehrlich, Michigan State University (ehrichn@msu.edu); Larry Hembroff, Michigan State University (hembroff@msu.edu); Karen Clark, Michigan State University (clarkk@msu.edu)

We recruited a panel of 525 MSU students who agreed to take several iterations of a survey detailing their preferences in the election to be held in 2008, beginning with 17 candidates for nominations by the two major parties in November 2007, followed by 10 waves up to and including the week after election day. In Wave 1 we collected demographics on age, race, sex, and enrollment and each respondent rated him/herself on three Likert scales of, social and fiscal ideology (conservative – liberal) and party identification. All the respondents were registered voters.

The impetus for the research was to document the degree to which individuals changed their preference, both from different candidates from the same party and across party lines.

Among the more notable results, we found

- respondents who were more polar in their party identification were least likely to make any change, even within their party, unless their candidate dropped out of the race, and very unlikely to cross party lines
- moderates favored Democrats 2:1 in November 2007 but voted Democratic 4:1 in 2008
- once Clinton dropped out of the race, no subsequent event precipitated any changes

From our analysis we concluded that

- no Republican could have beaten Obama
- no Republican would have lost to Clinton

Elections and Institutions: Ranked-Choice Voting and Its Attitudinal and Behavioral Effects.

Joanne M Miller, University of Minnesota (jomiller@umn.edu); Lawrence R Jacobs, University of Minnesota (ljacobs@umn.edu)

Institutions structure electoral behavior. The comparatively low turnout in the U.S. and depressed level of turnout among individuals with lower social and economic status has generated a series of reform proposals to increase turnout and widen participation among sub-groups that generally turn out at lower rates. This paper investigates the impacts of institutions on political behavior and attitudes through an unusual quasi-experiment in the Twin Cities in November of 2009 when municipal elections in Minneapolis used ranked-choice (or instant runoff) voting and the neighboring city of St. Paul used the traditional plurality-winner voting system. The fact that two cities less than 15 miles apart that are so demographically similar held city elections on the same day using two very different voting methods allows for a unique examination of political behavior and public opinion regarding ranked choice voting. Our study is based on an original RDD telephone survey of voters and non-voters in Minneapolis and St. Paul in the days following the November election. Our analysis focuses on the degree to which Minneapolis voters actually used the ranked choice system, comparisons between Minneapolis voters and non-voters in their evaluations of ranked-choice voting in 2009 and its possible use in the 2010 gubernatorial race (including their likelihood of voting in 2010), and variations in terms of demographic and attitudinal subgroups and individual-level resources (namely, political knowledge and organizational affiliations). A key element of our analysis will focus on the relative confidence of Minneapolis and St.

Paul voters that their votes were counted accurately and evaluations of the fairness of the election to gauge whether, and among whom, ranked-choice voting exhibited attitudinal effects. Our results have implications for the impact of ranked-choice voting and other electoral reforms on increasing voter turnout among low SES citizens and increasing the ideological diversity of candidates.

Questionnaire Design: IVR, Visual Scales, & Cognitive Interviewing.

Does Voice Matter for Youth Reports of Tobacco Use? An Interactive Voice Response Experiment

Niki Mayo, RTI International (nmayo@rti.org); Doug Currivan, RTI International (dcurrivan@rti.org); Brenna Muldavin, RTI International (bmuldavin@rti.org)

Household telephone surveys on youth tobacco use yield lower estimates than school-based self-administered surveys. Researchers generally assume the lower estimates from telephone surveys reflect underreporting due to youths' concern about parents or others overhearing their responses. Interactive voice response (IVR) has been shown to generally increase youth reports of smoking compared to interviewer-administered CATI surveys. Nevertheless, a significant gap remains between estimates from IVR and school-based surveys. One potential limitation of the IVR mode is that the "standard" human voice used to record the survey items is an adult female. Youth who respond via IVR may still feel like they are reporting to an adult, even though the voice is part of a computerized system. This may discourage some youths from fully reporting smoking behavior, as it could remind them that there is some risk, however small, of their smoking behavior being disclosed to their parents. For in-school surveys, the primary risk is that classmates who are also participating in the survey could see youths' responses.

In order to assess what role, if any, concerns about disclosure play in youth smoking reports in IVR, we designed a survey experiment where youth respondents were randomly assigned to one of two different voices – an adult female who sounded old enough to be the respondents' mother and a teen female who sounded close to the expected average age of respondents. To test whether the IVR voice had any impact on youth smoking reports, we compared responses across the two experimental conditions on three smoking behavior items. We also examined whether the IVR voice led to different response patterns for additional questions on youths' intentions to smoke. This paper reports findings from the experiment and discusses the implications of the findings for measuring youth smoking in telephone surveys.

Visual Effects: A Comparison of Visual Analog Scales in Models Predicting Behavior

Randall K. Thomas, ICF International (randall.k.thomas@gmail.com)

Multiple theories have linked attitudes, behavioral intention, and behavior. The Theory of Reasoned Action improved prediction of behavior by including three components: attitude, subjective norm, and intention. The Theory of Planned Behavior added behavioral difficulty to improve prediction. One unexplored area is how response formats can influence models of behavior. With the advent of computer-based surveys, the interactive visual scale (Visual Analog Scale – VAS) provided for potentially improved measurement precision and validity. While previous studies on the VAS compared scales on a singular basis and found that they were not better than simpler single response measures, this study examined their utility when multiple measures are used in models, since improved precision could yield improved differentiation across measures. Respondents participated in a web-based survey and were randomly assigned to evaluate 1 target behavior (out of 12 possible behaviors) and then asked about past 30 day behavioral frequency. They were then randomly assigned 1 of 5 response formats to evaluate 9 attitudinal components of attitude toward the behavior on a single screen. Two response formats were horizontal

radio button scales, 3 were Visual Analog Scales (each having a different presentational format). The horizontal scales took less time to complete than the visual analog scales. In addition, across the models predicting behavior, response format significantly affected item coefficients, demonstrating that choice of response format is a key aspect in comparing models and their efficacy.

Using Administrative Records to Evaluate Survey Responses in the National Immunization Survey

Ting Yan, NORC at University of Chicago (yan-ting@norc.org); Kennon Copeland, NORC at the University of Chicago (copeland-kennon@norc.org); Kirk Wolter, NORC at the University of Chicago (wolter-kirk@norc.org); Phil Smith, CDC (pzs6@cdc.gov)

Response error, one component of total survey error, can result in both increased variability and bias in survey estimates. Administrative records provide a source for evaluating response error. The National Immunization Survey (NIS) – a nationwide, list-assisted RDD survey sponsored by the Centers for Disease Control and Prevention and fielded by NORC – monitors the vaccination rates of children ages 19-35 months and adolescents age 13-17 years. Data collection for the NIS involves both a household interview and a provider survey, each of which is subject to response error. One objective of the NIS-Registry project was designed to examine potential coverage, response, and nonresponse error in the NIS. A sample of NIS-eligible children was selected from the state immunization information systems (IIS) in Arizona and Michigan. The NIS household and provider-record-check (PRC) questionnaires were administered for the samples during Q1/2008 and Q2/2008.

This paper focuses on the set of children from the IIS samples for whom NIS household and PRC data are available. These data are compared to each other (thus treating the PRC as administrative data) and to the IIS data (thus treating the IIS as administrative data and both the household and the PRC as response data). We conduct the analyses under the total survey error paradigm. The purpose is to identify respondent characteristics associated with response error. In addition to sociodemographic characteristics, we examine respondents based on contactability and level of cooperation to study their response errors.

Questionnaire Design Considerations when Expanding a Survey Target Population to Include Children

Patricia LeBaron, RTI International (plebaron@rti.org); Rebecca Granger, RTI International (rebecca@rti.org)

When designing a questionnaire, survey practitioners are challenged with developing questions that are appropriate for all respondents within the target population. This task becomes increasingly difficult when the same survey instrument will be administered to both children and adults.

This presentation will examine factors to be considered when asking children to respond to the same survey items asked of adults. The authors will present a review of the literature, citing cognitive difficulties that younger respondents encounter when reporting past behaviors. Questions may have to be reworded to allow younger respondents to comprehend questions within a given questionnaire. Survey designers can also expect younger respondents to have greater difficulty recalling past behaviors, hindering their ability to make reports on these behaviors.

In order to evaluate the impact of expanding the target population of an existing survey on questionnaire design and unit and item nonresponse, the authors will examine data from the National Survey on Drug Use and Health (NSDUH). The NSDUH is an annual face-to-face, household survey that interviews people 12 years old and older sponsored by the Substance Abuse and Mental Health Services

Administration and is the nation's leading source of information on substance use behaviors. We will examine differences in item missing data, unit nonresponse and timing data between minors in the sample. This analysis will show how the response behavior of 12 year olds differs from 13 year olds, and how these differ from 14 year olds, for example. The differences in data quality and respondent behavior between these two age groups will provide insight into the effects of administering a survey to respondents under 12. This presentation will provide suggestions to questionnaire designers that are either designing new questionnaires for younger populations or those that are amending existing instruments to be administered to younger respondents.

Mixed Modes and Measurement Error: Using Cognitive Interviewing to Explore the Results of a Mixed Modes Experiment

Pamela C. Campanelli, The Survey Coach (pamc@aspects.net); Margaret Blake, National Centre for Social Research (Margaret.Blake@natcen.ac.uk); Michelle Gray, National Centre for Social Research (Michelle.Gray@natcen.ac.uk); Steven Hope, National Centre for Social Research (Steven.Hope@natcen.ac.uk); Annette Jackle, University of Essex (aejack@essex.ac.uk); Peter Lynn, University of Essex (plynn@essex.ac.uk); Alita Nandi, University of Essex (anandi@essex.ac.uk); Gerry Nicolaas, National Centre for Social Research (Gerry.Nicolaas@natcen.ac.uk)

A three-year project to inform decisions about when and how to mix modes was launched in Great Britain in October 2007 as part of an ESRC initiative. Part of the project focused on the collection of new experimental data. Survey participants from an earlier phase of the project who had access to the internet were randomly assigned to one of three modes: CAPI, CATI and CAWI. The questionnaire included 67 questions purposely chosen or newly written to test hypotheses based on question task difficulty, question sensitivity, question type (including satisfaction, other attitudinal, behavioral and other factual) and question format (including short versus long scales, rating versus ranking, agree/disagree statements, "yes/no for each" versus "mark all that apply", branching versus no branching, fully-labeled scales versus end-labeled scales and showcard versus no showcard on long lists in CAPI). The initial quantitative analysis showed some hypotheses supported, some not supported and some surprising findings. The project had been designed to have a post analysis cognitive interviewing phase to provide deeper understanding of the quantitative results. All of the surprising findings as well as a selection of other findings that deserved further investigation were selected for the cognitive interviewing phase. Thirty six respondents were recruited from those who had participated in the mixed modes survey using specific quotas, contrasting respondents who displayed satisficing behaviour versus those who did not. In the first part of the cognitive interview each respondent was asked questions in three different modes (CATI, CAWI and CAPI). This was followed by retrospective think alouds and probing. All cognitive interviews were then transcribed and the data introduced into the qualitative charting programme, "Framework". This paper discusses the highlights of what the cognitive interviewing discovered and how different strategies of cognitive interviewing had to be employed.

Rethinking Response Rates

Rethinking Response Rate Targets: New Evidence of Little Relationship Between Response Rate and Observable Response Bias

Richard Hendra, MDRC (richard.hendra@mdrc.org); Aaron Hill, MDRC (aaron.hill@mdrc.org)

A flurry of recent research has suggested that increased response rate often does little or nothing to diminish the presence of observable response bias. Building on this knowledge, the authors analyzed a rich dataset of 16 comparable surveys with nearly 9,000 responses that represent more than 27,000 research subjects. The authors used detailed disposition data to simulate response rates from 40% to over 80% within each survey, measuring and comparing the levels of response bias. This yielded a database of 140 survey samples to evaluate. The analysis found little relationship between the rate of response and presence of response bias.

This paper adds important new information and context to the discussion about response rate targets. The authors had access to micro-data, which made comparisons between samples more precise. They had extensive data on the full sample, including all nonrespondents, such as demographic characteristics and data from administrative records. The surveys, although administered at different sites, were very similar in nature and were administered in the same manner (a mixture of telephone and in-person interviews). The surveys were conducted within the framework of experimental, randomized controlled trials; thus, the authors are able to demonstrate the relationship of differences in background characteristics between experimental and research groups as response rate changes.

The database is part of the Employment Retention and Advancement (ERA) project, a national project that encompasses more than a dozen demonstration programs to discover what approaches help welfare recipients and other low-income people stay steadily employed and advance in their jobs.

As researchers from all fields are experiencing a decline in survey response rates, many practitioners have begun to question the benefit of pushing to reach higher rates of response. The results of this paper add further evidence that extremely high response rates do not necessarily reduce response bias.

Using Imputation Methods to Estimate “e”

James Wagner, University of Michigan/Survey Research Center (jameswag@isr.umich.edu)

AAPOR defines response rates that include an adjustment factor for cases that have unknown eligibility at the end of the survey. This factor is known as “e”. Typically, people use the eligibility rate from the part of the sample where the eligibility status (either eligible or ineligible) is observed. This estimate is sometimes called the CASRO estimate of e.

However, in a telephone survey, this estimate of e is likely to be biased upwards for the unknown part of the sample. Many of the cases that are never contacted are not households. They are simply numbers that will ring when dialed, but are not assigned to a household. None of these cases are ever involved in estimates of e. This leads to the biased estimate.

Brick and Montaquila (POQ, 2002) described an alternative method of estimating e. They use a survival model. This lowers estimates of e relative to the CASRO method. But it is still upwardly biased since many of the cases that are not screened for eligibility could never be contacted since they are not actually households.

This paper will propose an alternative method to the survival model. The proposed method is to use imputation methods to estimate the eligibility among the unknown cases. Although the methods are similar, the imputation approach has several advantages. First, it develops a range of estimates. This is a useful reminder of our uncertainty. Second, it allows great flexibility. It is very easy to include covariates in the model. It is not as easy to include covariates in the survival model. Finally, it introduces the possibility

of specifying a model for this situation that addresses the nonignorable nature of the missing data problem.

Participation in Surveys in the U.S.: A Three Year Study.

David Dutwin, Social Science Research Solutions (ddutwin@icrsurvey.com); Robyn Rapoport, SSRS (rrapoport@ssrs.com)

The ICR/SSRS omnibus survey EXCEL is one of the longest running surveys in the nation. It has, in fact, run at least weekly for nearly 25 years, gathering on average over 60,000 interviews on an annual basis. The survey presently utilizes a 6 call rule design for 1,000 interviews per wave. Nearly one million sample records are generated annually for the survey. This paper presents findings from an analysis of the call dispositions of nearly 3 million sample records, whose outcome measure is the likelihood to fall into one of the four main AAPOR response rate categories (complete, non-interview but eligible, not-eligible, and unknown if eligible). Utilizing MSG-Genesys data at the telephone exchange level, we analyze AAPOR outcomes by racial, ethnic, income, education, and age densities, as well as by state, and metropolitan status. Likelihood to complete by number of call attempts is considered as well. Interactive effects are reported in addition to main effects. The results show strong differences by a wide range of variables to participation, and provide implications for sample needs in regionally limited studies and considerations for non-response adjustments to weighting.

An Alternative to the Response Rate for Measuring a Survey's Realization of the Target Population

Benjamin Skalland, NORC at the University of Chicago (skalland-benjamin@norc.org); Martin Barron, NORC at the University of Chicago (barron-martin@norc.org); Ashley Amaya, NORC at the University of Chicago (amaya-ashley@norc.org)

A survey's response rate is the number eligible units with a completed interview expressed as a proportion of the estimated number of eligible units in the sample. Rightly or wrongly, the response rate is often taken to be the primary (or sole) measure of survey quality when assessing the validity of survey data or comparing between different surveys. However, the response rate has three major limitations: (1) the calculation of the response rate is very sensitive to the choice of "e", the assumed proportion of eligible units among the non-responding units; (2) the response rate is based on the sample, not the population, and so does not account for the sampling frame's under-coverage of the eligible population; (3) the response rate does not account for misclassification of eligible units as ineligible. Because of these limitations, the response rate may be inadequate for comparing different surveys' ability to identify and complete interviews for the target population. In recent years, this inadequacy has become more apparent with the increase in sampling frame options with differing and often unknown coverage properties (landline RDD, list-assisted landline RDD, landline + cell RDD, address-based sampling) and the increase in the use of data collection modes with differing non-response mechanisms (telephone, mail, web, face-to-face). In this paper, we describe a benchmarked "realization" rate – a frame-independent, assumption-free measure of a survey's ability to identify and interview the eligible population, based on an external measure of the size of the eligible population. We argue that the benchmarked realization rate is superior to the response rate for comparing surveys that use different sampling frames or modes of data collection. We provide an example calculation and discuss the advantages and disadvantages of this approach.

Sampling Hard-to-Reach Populations

Using Lists with RDD Samples: An Examination of Bias, Cost and Variance Estimates.

Anthony M Roman, Center for Survey Research UMass-Boston (anthony.roman@umb.edu); Charles F Turner, City University of New York, Queens College and Graduate Center (CFTurner@popepi.org); Susan M Rogers, Research Triangle Institute (smr@rti.org); Elizabeth Eggleston, Research Triangle Institute (eeggleson@rti.org); James R Chromy, Research Triangle Institute (jrc@rti.org); Sylvia Tan, Research Triangle Institute (stan@rti.org)

The Monitoring STIs Survey Program (MSSP) was a telephone survey conducted in Baltimore, Maryland from 2006 through 2009. The project screened households for people 15-35 years of age and then randomly selected one eligible person for interview. Those that completed an interview were then asked if they would submit a urine sample for testing if mailed a specimen cup. Primary analytic variables of interest included whether these tests indicated the presence of any of three sexually transmitted infections (STIs). To gain efficiency and reduce costs, early in the study a methodological change was implemented. The sampling frame for the project was altered from a pure random-digit-dialed (RDD) sample to one which incorporated the use of lists that identified the ages of residents of some households. This paper will examine the potential biases that may be present in the use of lists alone. It will also address the tradeoff of increased sample size for a fixed survey cost compared to increased variance estimates due to a more complex sample design.

Screening for Specific Population Groups in Mail Surveys

Douglas Williams, Westat (douglaswilliams@westat.com); Jill M. Montaquila, Westat (jillmontaquila@westat.com); J. Michael Brick, Westat (mikebrick@westat.com); Mary C. Hagedorn, Westat (maryhagedorn@westat.com)

The National Household Education Surveys Program (NHES) has used RDD/CATI methodology to survey households since 1991. NHES is the only household-based source of national data on topics ranging from early childhood care and school readiness to the activities of older school age children.

To address challenges in the current survey climate, NHES conducted a number of experiments with a variety of design alternatives. The greatest single design change was the movement from RDD and telephone administration to mail administration of the screener and topical surveys. To determine the feasibility of the new design, a pilot study was conducted. The pilot consisted of a sample of 11,800 households selected from an address based sample using the USPS Delivery Sequence File. Data collection for the pilot used a two-phase approach: (1) all households were mailed a screening questionnaire; (2) households with eligible children (identified in the screener) were mailed a topical questionnaire.

Implications for screening for specific population groups (households with children) are discussed by analyzing the results of manipulating the screening request. Different versions of the screener package that tailor the survey request reveal differences in overall response. We will show how these differences extend to population groups of interest. Approaches used and issues considered in the design of the different questionnaires and materials will also be discussed. Finally we will show comparisons of the effectiveness of the different approaches for a national sample compared to a targeted sample of households with children, and discuss the implications of the findings for a survey that targets households with children.

An Oversampling Method for Youth in a Smoking Survey

Peter Wright, Statistics Canada (peter.wright@statcan.gc.ca); Benjamin Hartling, Statistics Canada (benjamin.hartling@statcan.gc.ca)

While smoking cessation methods and products are targeted to older cigarette smokers, younger people are targeted for campaigns to prevent and discourage them from developing smoking habits. To conduct a random-digit dialing survey of smoking habits and attitudes about smoking, we oversample younger people in order to improve the estimates associated with their particular smoking habits and attitudes. This method is applied instantaneously as the household roster is completed by computer-assisted telephone interviewers for a national tobacco use monitoring survey. The method involves pre-generating a set of random numbers to apply to the household, then a set of deterministic rules determines the persons in the household to select for interviews. We discuss extending this approach in order to target any subgroup based on the roster information, along with its advantages and limitations.

A Comparison of Black Oversamples: Census Density vs. Call-Backs

Peyton M. Craighill, ABC News (peyton.m.craighill@abc.com); Gary Langer, ABC News (gary.langer@abc.com); Patrick Moynihan, Harvard University (moynihan.patrick@gmail.com); Melissa Herrmann, SSRS (mherrmann@ssrs.com); David Dutwin, SSRS (ddutwin@ssrs.com)

Attitudes of African-Americans are of special interest in a range of public opinion polls, but the population of blacks is insufficient for an adequate sample size in customary national RDD surveys of 1,000 adults. Two oversampling methods are common: One, employing callbacks to known African-American households contacted in previous surveys; and two, calling into Census tracts with disproportionately high concentrations of African-American households. ABC News, working with the research firm SSRS, tested both approaches in a survey focused on black election politics, co-sponsored by Columbia University and USA Today. This paper compares the attitudinal and demographic differences between the two samples, with further comparisons to Census parameters and to RDD data from contemporaneous ABC News/Washington Post surveys. We also evaluate the efficiency and comparative costs of each method.

Confirmation Bias in Housing Unit Listing

Stephanie Eckman, University of Maryland (steph.eckman@gmail.com)

Field listing of housing units is an expensive and time-consuming stage of the survey process, and its error properties are not well understood. This paper uses an experimental repeated listing design to demonstrate the presence of confirmation bias in dependent listing. When provided with an initial listing to update in the field, listers become too trusting of the list and tend not to add missing units or delete inappropriate units. This finding has implications not only for surveys that use dependent listing to create housing unit frames but also for studies of coverage of housing unit frames.

Surveys on Medical & Health Issues

Health Care Status over Time: Lessons from a Follow-Back Survey of Individuals with Special Health Care Needs

Shannon R. TenBroeck, NORC at the University of Chicago (tenbroeck-shannon@norc.org); Kathleen S. O'Connor, National Center for Health Statistics (NCHS), CDC (kdo7@cdc.gov)

Often survey data offer a snapshot of reality. However, a follow-back design, such as the one employed by the 2007 Survey of Adult Transition and Health (SATH), provides a unique opportunity to study changes in health care over time and the effect these shifts may have on health status.

For this paper, we will examine data from the 2001 National Survey of Children with Special Health Care Needs (NS-CSHCN) and the 2007 SATH. Both surveys were conducted through the State and Local Area Integrated Telephone Survey (SLAITS) mechanism of the National Center for Health Statistics (NCHS) on behalf of the Maternal and Child Health Bureau. In the 2001 random digit dial (RDD) survey, data were collected from parents and guardians of children with special health care needs (CSHCN). The survey collected information on the selected CSHCN's medical home, health insurance coverage, and access to care. For the SATH follow-back survey, young adults (ages 19-23) – the subjects of the NS-CSHCN interview six years earlier – were contacted in an effort to shed light on their updated health care needs.

Our analysis will involve the evaluation of three groups: SATH respondents who report that, since 2001, their general health has a) improved, b) declined, or c) remained the same. Because identical questions were asked of the parent/guardian in 2001 as were asked of the young adult in 2007, we are able to explore which aspects of a subject's health care changed, and in what direction. Despite the change of respondents, meaningful conclusions can be drawn due to the objective nature of factors we will examine. Specifically, we will focus on health care-related factors, such as medical home, health insurance coverage, and access to care. We will also explore indicators of financial stability such as income and ability to pay for needed services.

Respondents' Methods for Managing and Recalling Data in the Medical Expenditure Panel Survey

Sid J Schneider, Westat (schneis1@westat.com); Barbara Gustavson, Westat (BarbaraGustavson@westat.com); Tiandong Li, Westat (TiandongLi@westat.com); Huseyin Goksel, Westat (HuseyinGoksel@westat.com); Ralph DiGaetano, Westat (RalphDigaetano@westat.com); W. Sherman Edwards, Westat (ShermEdwards@westat.com)

This study investigated respondents' methods for managing and recalling personal health data for a household survey.

In the Medical Expenditure Panel Survey - Household Component, interviewers visit a nationally representative sample of households five times over two years. They collect data about each resident, including demographic background, income, insurance coverage, and every health-related product and service that each resident has used, including its purpose, date, provider, and cost. From 2004 to 2006, 29,798 families participated, reporting 667,094 medical products and services.

The interviewers observe the methods that the respondents employ to recall their families' use of health-related products and services. Those methods are categorized as memory, checkbook, calendar, prescription container label, documentation such as bills and receipts, and other. Interviewers could record that respondents used one or more than one method to recall data about a medical product or service. All respondents are given calendars with pockets for storing documentation.

Logistic regression analyses were performed to investigate the variables that were associated with the respondents' recall methods. The results suggest that variables that may be proxies for low socioeconomic status—lower family income, less educational attainment, Medicaid coverage, and lack of private health insurance—were associated with less use of calendars, checkbooks and documentation, and greater use of prescription labels and memory. Younger, male, Hispanic and non-Caucasian respondents also exhibited that pattern.

The respondents' geographic location and the nature of the product or service that the respondents reported also affected the recall methods. The implications of the findings for health-related household surveys will be discussed.

This research was conducted by Westat and Insight Policy Research under contract to the Agency for Healthcare Research and Quality, Contract HHSA290200710072T. The authors of this presentation are responsible for its content. No statement may be construed as the official position of AHRQ or the Department of Health and Human Services.

Use of Household Immunization Records (Shot Cards) on the National Immunization Survey
Ashley Amaya, NORC at University of Chicago (amaya-ashley@norc.org); Xian Tao, NORC at the University of Chicago (tao-xian@norc.org); Karen Wooten, Centers for Disease Control and Prevention (kgw1@cdc.gov)

The National Immunization Survey (NIS) is a nation-wide list-assisted RDD survey which seeks to collect childhood immunization histories from approximately 27,000 children per year. As part of the interview, the NIS requests the caregiver retrieve and refer to the child's vaccination records or shots, if available. Regardless of shot card availability, the NIS also asks for permission to contact the child's immunization provider(s) and request vaccination information (dosage, type, and date) directly from these providers.

In this paper we review the NIS methodology for requesting and referencing vaccination records. We discuss different methods of requesting records' retrieval, the proportion of people who retrieve records for the interview, differences between respondents who do and do not retrieve records for the interview, and the completeness of information from those records compared to vaccination records from immunization providers.

Internet Panels and Health Research: Findings from National RDD Surveys.
John M. Boyle, Abt SRBI Inc. (j.boyle@srbi.com)

Internet panels are widely used in market research because they can provide samples of relatively low incidence subpopulations quickly and at low cost. Critics argue these are non-probability samples with known and unknown biases. Proponents argue that known demographic biases are corrected by sample balancing and weighting. This study examines the potential usability of Internet panels for non-commercial health research.

As part of a study of asthma and its management in the United States, two national surveys were conducted by telephone in 2009. The first was a national probability sample of 2,500 persons with current asthma, obtained by telephone screening of 60,000 households sampled by random digit dialing. The second was a national sample of 1,090 adults sampled by RDD. In both samples, the respondents were asked whether they participate in any internet panel surveys, which ones, and how often they are contacted to participate in Internet surveys.

If Internet survey panelists are defined as persons who belong to panels and are contacted at least once a month to participate in them, then the national RDD surveys suggests approximately five percent of the adult population of the United States are Internet panelists. These Internet panelists differ from those who do not belong to Internet panels on a number of key demographic characteristics including age, income, race and ethnicity, marital status, employment status and housing type.

Despite known demographic biases, can Internet panels provide representative samples of relatively rare disease populations, like asthma? This paper compares the disease and treatment characteristics of nearly two hundred Internet panelists with more than two thousand non-panelists from the national RDD asthma patient survey. We find symptom and treatment characteristics are similar for panel and non-panel members with asthma. Hence, Internet panels should not be dismissed out of hand as a potential source for health research.

Assessing the Accuracy of the Face-to-Face Recruited Internet Survey Platform: A Comparison of Behavioral and Health-Related Findings from the FFRISP with Those of Major National Surveys
Ana Villar, Stanford University (villar@stanford.edu); Ariel Malka, Yeshiva University (amalka@yu.edu); Jon Krosnick, Stanford University (krosnick@stanford.edu)

The Face-to-Face Recruited Internet Survey Platform (FFRISP) is an internet survey panel that attempted to collect information that is representative of the general American population, while drawing on the advantages of working with such a panel, namely, obtaining data more quickly, more accurately, and less costly than using interviewer administered survey modes.

The goal of the research described in this presentation is to examine the accuracy of the FFRISP by comparing survey estimates generated from it to those of the National Health and Nutrition examination supplement (NHANES) and from supplement surveys of the Current Population Survey (CPS). Survey data from the CPS and the NHANES have been used as benchmarks for behavioral data. Both surveys are conducted through interviewer-administered surveys on very large samples of individuals, with high response rate outcomes.

The paper starts by comparing distributions of questions about health, physical activities, and mood disorders—such as general self-perceived health, biking to work, or depression feelings—across the NHANES and FFRISP surveys. It then compares distributions of identically worded items about shopping behavior across the CPS Food Security supplement and FFRISP surveys. These comparisons will be made looking at full sample distributions, as well as at distributions of subsamples (e.g., respondents with lower education vs. respondents with higher education). The paper will then compare whether associations between variables are similar across these surveys. Similarities and differences between the findings generated with FFRISP data and those generated with CPS and NHANES data will be discussed.

AAPOR Concurrent Session A

Sunday, May 16, 2010

8:30 a.m. - 10:00 a.m.

Design Considerations for Establishment Surveys

Nonresponse Reduction Strategies in Establishment Surveys: Design Considerations in the National Survey of Residential Care Facilities

Lauren Harris-Kojetin, National Center for Health Statistics, CDC (fti3@cdc.gov); Tim Flanigan, RTI International (tsf@rti.org); Angela Greene, RTI International (amg@rti.org); John Loft, RTI International (jloft@rti.org); Abigail Moss, National Center for Health Statistics (ajm1@cdc.gov); Eunice Park-Lee, National Center for Health Statistics (hta8@cdc.gov); Emily Rosenoff, U.S. DHHS, Office of Assistant Secretary for Planning and Evaluation (Emily.Rosenoff@HHS.GOV); Manisha Sengupta, National Center for Health Statistics (hku2@cdc.gov); Sara Zuckerbraun, RTI International (szuckerbraun@rti.org)

Obtaining higher response rates is becoming increasingly difficult in both household and establishment surveys that provide critical information for policymaking and planning. There are cooperation-related considerations unique to surveying establishments. For example, gatekeepers may pose challenges in gaining access to appropriate respondents to obtain cooperation; whether to participate in a survey may be influenced by organizational challenges (e.g., unavailable staff) and priorities; complex corporate structures may need to be navigated to obtain cooperation; and, multiple respondents or sources may be needed to provide the requested information. Differences between responding and nonresponding establishments could have a greater effect on bias than in household surveys when establishments vary greatly in size and structure. This presentation highlights multiple strategies to be used before and during data collection to minimize unit nonresponse in the first ever National Survey of Residential Care Facilities (NSRCF), an in-person establishment survey scheduled to be fielded March through September, 2010. The U.S. Department of Health and Human Services (DHHS) is conducting NSRCF. Nonresponse reduction strategies discussed in the presentation are based on lessons learned from the NSRCF pilot and pretest, completed in 2008, and build on experiences conducting other DHHS-sponsored health care establishment surveys. Strategies include promotion and outreach to industry associations; letters of support from industry associations; outreach to corporate headquarters of sampled establishments; advance materials and use of multiple contact modes to reach respondents; use of recruiters with refusal avoidance experience; emphasis on flexibility in meeting facility staff availability and schedules; mid-field debriefings with recruiters to identify and address challenges to obtaining cooperation; and, tailored non-response follow-up letters. This presentation is intended to provide practical applications in designing other surveys and to promote discussion about ways to enhance the evidence base for nonresponse reduction strategies in establishment surveys.

Measurement Issues in Establishment Surveys: Design Considerations in the National Survey of Residential Care Facilities

John D Loft, RTI International (jloft@rti.org); Tim Flanigan, RTI International (tsf@rti.org); Angela Greene, RTI International (amg@rti.org); Lauren Harris-Kojetin, National Center for Health Statistics (fti@cdc.gov); Abigail Moss, National Center for Health Statistics (ajm1@cdc.gov); Eunice Park-Lee, National Center for Health Statistics (hta8@cdc.gov); Emily Rosenoff, U.S.

DHHS, Office of Assistant Secretary (Emily.Rosenoff@hhs.gov); Manisha Sengupta, National Center for Health Statistics (hku2@cdc.gov); Sara Zuckerbraun, RTI International (szuckerbraun@rti.org)

Establishment surveys present challenges in the collection of high quality data that are distinct from population-based surveys. Although cognitive models of survey response developed in the context of household surveys are valuable tools, the complexity of establishments as data collection units suggests additional factors for consideration in survey design. For example, an establishment's size and business practices related to the location of and control over information about the establishment may mean that several individuals within the establishment are appropriate survey respondents. Data retrieval tasks undertaken by respondents within an establishment might require recalling from memory or abstracting from administrative records. Interpretations of terms and nomenclature used in questionnaires might vary across establishments due to variation in geographic area, size, or type of facility. Another measurement issue relevant to establishment surveys is the use of sampling frames of establishments as potentially valuable sources of data to link with survey data. Sampling frame data may even be used to define sampling strata. However, frame data may be inconsistent with establishment-provided responses and, therefore, need to be verified by the establishments during the course of survey data collection. This paper discusses these measurement considerations in the context of planning for the first ever National Survey of Residential Care Facilities (NSRCF), an establishment survey conducted by the U.S. Department of Health and Human Services. The NSRCF pretest serves as a case study to provide concrete examples of conceptual issues. Data quality concerns in the specific context of these establishments led to approaches and tools to facilitate identifying appropriate respondents within the facilities, promote accurate data retrieval, and provide flexibility in managing data collection activities in facilities. Conclusions and generalizations are offered to assist in the design of establishment surveys and encourage research on related methodological issues.

Toward a Standard Typology of Disposition Codes for Establishment Surveys

Sara Zuckerbraun, RTI (szuckerbraun@rti.org); Laura Flicker, RTI (lflicker@rti.org); Jamie Friedman, RTI (jfriedman@rti.org)

Disposition codes are used to track the status of cases during data collection. The AAPOR Standard Definitions (1) provide guidance on establishing final disposition codes, converting from pending to final codes, and computing final response rates for household, mail and internet surveys. However, there is less information publicly available about establishing disposition codes and calculating response rates for establishment surveys. The authors are survey researchers engaged in large-scale federal government surveys of establishments in the health care and education industries. We suggest that there are categories of pending and final dispositions that are likely needed by many if not most establishment surveys. Some of these categories are comparable to the standard household survey codes, but many reflect the different nature of establishment surveys and do not have a household counterpart. We suggest eight categories of disposition codes: 1) sample integrity, 2) identification of the target respondent 3) sending materials to the establishment, 4) gaining the approval of the target respondent, 5) gaining the approval of other entities apart from the target respondent and/or outside the establishment, 6) setting appointments, reminding, and prompting 7) post-interview tasks including validation, verification, and retrieval, and 8) post-data collection "thank you's" to the establishment. Because establishment surveys tend to have complex designs and multiple phases of data collection, and often collect data from multiple respondents and different instruments, we suggest a structure of disposition codes that captures the latest event separately from the phase(s) and status. We also give advice about how to "roll up" phases to the general case level in order to calculate response rates and other performance rates of interest.

The typologies presented will be useful to survey researchers and practitioners in designing the case management system and in understanding and calculating response rates of establishment surveys.

Meeting the Challenges of Converting a Large Establishment Survey from Paper to Electronic Administration

David James Roe, RTI International (droe@rti.org); Lisa Thalji, RTI International (thalji@rti.org); John Loft, RTI International (jloft@rti.org); Laura Flicker, RTI International (lflicker@rti.org); Jason Stockdale, RTI International (jstockdale@rti.org); Marie Stagnitti, Agency for Healthcare Research and Quality (AHRQ) (Marie.Stagnitti@ahrq.hhs.gov)

As technological advances continue to foster the move from paper based survey administration to computer based administration, consideration must be given to the potential effect that switching modes may have on a given questionnaire. Issues related to mode and method effect, usability issues, adjustments to question wording and response formats are well documented. However, less documentation exists on the operational considerations for converting an entire study protocol from paper-based to computer-based administration. These are particularly complex issues for establishment surveys, where issues of case ownership, building consistent rapport with respondents, and managing multiple points of contact present additional challenges beyond those that are common to household and individual surveys. Sponsored by the US Department of Health and Human Services' Agency for Healthcare Research and Quality (AHRQ), the Medical Expenditure Panel Survey (MEPS) is a set of large-scale surveys of families and individuals, their medical providers, and employers across the United States. MEPS is the most complete source of data on the cost and use of health care and health insurance coverage. In 2010, the Medical Provider Component (MPC) of MEPS, which covers hospitals, physicians, home health care providers, and pharmacies identified by respondents in the MEPS Household Component, will undergo a major change from paper based administration to computer based administration. This change not only affects the way data will be collected via instrumentation, but also how sample records, case level details, points of contact, problems during data collection, issues of nonresponse and interviewer productivity are managed and tracked. This presentation will share information on some of the challenges presented by such a conversion, along with the unique and customized solutions developed in order to prepare the MEPS-MPC for data collection in 2010.

AAPOR Concurrent Session A

Sunday, May 16, 2010

8:30 a.m. - 10:00 a.m.

Assessing Data Quality: Paradata, Monitoring & Editing

The Accuracy of Interview Paradata: Results from a Field Investigation. Paul Biemer, RTI International and UNC-CH (ppb@rti.org); Kevin Wang, RTI International (kwang@rti.org)

Survey statisticians and methodologists are inventing new ways of using interview paradata. Novel uses range from monitoring and controlling errors in the field to post-survey adjustments. As an example, the authors are experimenting with using number of callback attempts and call outcomes for unit nonresponse adjustment. However, some problems with the quality of the paradata have limited the effectiveness of these uses. This paper reports on an investigation of paradata quality in a large-scale field survey. The study included both focus groups conducted with field supervisors and an informal survey of 601 interviewers regarding their practices in recording call attempts and call outcomes. A particular focus of this study is how interviewers record situations that are not be straightforward such as drive-by sightings, calls to neighboring units, encounters with gate keepers, and so on. The investigation identified a number of situations where call attempts and outcomes were recorded in error and where the collection of paradata was done in a manner inconsistent with the concept of “level of effort” that nonresponse adjustment models rely on. We speculate on the causes of the problems and discuss what steps could be taken to reduce the errors.

Who’s Monitoring the Monitors?: Examining Monitors’ Accuracy, Objectivity, and Consistency to Improve the Quality Assurance Process

Claudia Gentile, Mathematica Policy Research (CGentile@mathematica-mpr.com); Jason Markesich, Mathematica Policy Research (JMarkesich@mathematica-mpr.com); Joe Baker, Mathematica Policy Research (JBaker@mathematica-mpr.com); Shawn Marsh, Mathematica Policy Research (SMarsh@mathematica-mpr.com)

Interviewer monitoring is used by the majority of survey organizations to evaluate interviewer performance, identify problems with survey questions, and provide data for methodological investigations of questionnaire designs or interviewing techniques. Although monitoring is viewed as a key quality assurance tool, little research has been devoted to understanding the behavior of monitors—specifically, their ability to provide effective and consistent feedback on interviewer performance. To explore monitor behavior, Mathematica designed a “monitoring consistency” training exercise that addressed the following questions: (1) What are the typical behavioral issues that monitors focus on when evaluating interviews? (2) What criteria do monitors use to rate interviews? (3) Are monitors coding non-standardized interviewing behavior in an accurate, objective, and consistent manner? (4) What is the extent of “monitor variation,” within each monitor (drift across monitoring sessions) and between monitors (differences in leniency and severity between different monitors)? As part of this training, monitors were asked to evaluate tape-recorded interviews. To examine inter-rater reliability, we tabulated monitors’ ratings and evaluation feedback for each recorded interview. We compared the ratings to a criterion (gold standard) set by senior survey supervisors who were independent of the study, and also compared ratings among the monitors. To explore the ways monitors use criteria to assign ratings, we conducted think-aloud

protocols. While reviewing and rating tape-recorded interviews, monitors articulated their processes and decision-making, revealing how they apply the criteria when assigning ratings and providing feedback. We used findings from the consistency exercise, and the think-aloud protocols to: (1) develop training modules to improve the ways that monitors evaluate and provide constructive feedback to our telephone interviewers and (2) revise and improve our monitoring system, including the monitoring procedures, codes and form.

Standardizing Paradata

David Chearo, University of Chicago Survey Lab (dchearo@uchicago.edu); Martha Van Haitsma, University of Chicago Survey Lab (mvh@uchicago.edu)

The recording and analysis of paradata, data describing collection processes, are central to our development of methodological standards and best practices. While references to paradata have become commonplace in the world of public opinion research, the nature of the underlying data often remains ambiguous. This paper is largely a theoretical discussion about the need and possible strategies for the standardization of paradata.

After briefly defining the concept of paradata and situating it in earlier literature, we argue that standard documentation of data collection processes is needed to promote sharing and accurate comparison of methodological results. We outline some areas in which paradata could be standardized, highlighting those that should be the easiest and most important to implement uniformly across many research projects and institutions.

As part of the project of standardization, we propose a universal system of attempt codes. This system must be flexible and applicable to new and innovative research designs, many of which combine multiple modes and sample frames. To facilitate discussion, we present a specific example of such a system and demonstrate how its design provides an efficient medium for the development and testing of methodological advances.

Assessing the Quality of Survey Data

Joerg Blasius, University of Bonn (jblasius@uni-bonn.de)

Responses to a set of items in survey data are usually associated with socio-demographic characteristics such as age, gender, and educational level, but they are also associated with different kinds of response styles, such as acquiescence response style, extreme response style, and mid-point responding. Further, in different countries the meanings of words, phrases, or sentences may differ even when the translation is of high quality as it yields for the International Social Survey Program (ISSP). When analyzing a battery of items, responses are related to the substantive concept and to methodological effects. Applying multiple correspondence analysis and categorical principal component analysis to an item battery of survey data allows us to assess what part of the responses is due to substantive relationships and what part is attributable to methodological artifacts. Using different examples of ISSP data, we will show which part of variation can be explained by the substantive concept and which part is due to methodological induced variation.

How Much Does Past Experience Matter? An Investigation of the Factors Affecting Editing Quality and Efficiency

Geraldine Mooney, Mathematica Policy Research (gmooney@mathematica-mpr.com); Tedde Tsang, Mathematica Policy Research (ttsang@mathematica-mpr.com); Melissa Krakowiecki, Mathematica Policy Research (mkrakowiecki@mathematica-mpr.com); Cathie E Alderks,

**Substance Abuse and Mental Health Services Administration (cathie.alderks@samhsa.hhs.gov);
Deborah Trunzo, Substance Abuse and Mental Health Services Administration
(deborah.trunzo@samhsa.hhs.gov)**

When surveys include paper questionnaires, editing completed paper questionnaires prior to data entry helps minimize errors, facilitate smoother data processing, and increase confidence in the final data. To increase efficiency and accuracy, survey researchers often prefer to use subject-matter specialists as editors because of their ability to understand survey content (Granquist 1995; EUROSTAT 2007). Consequently, surveys often pay a premium to employ a high proportion of subject-matter specialists to perform editing. Similarly, the National Survey of Substance Abuse Treatment Services (N-SSATS), a multi-mode (web/mail/CATI) annual survey of approximately 17,000 substance abuse treatment facilities, has traditionally placed a priority on rehiring high-performing N-SSATS editors from previous years. Despite this long-standing preference, little research has been conducted on whether paying a premium for past experience is worth the additional investment or whether a well planned and executed editor training can sufficiently compensate for inexperience.

This paper will examine data editing records from the past five years (2005-2009) of N-SSATS in order to investigate: (1) whether past N-SSATS editing experience results in higher quality and/or more efficient editing and (2) what overall factors contribute to higher quality and more efficient editing. To investigate the former, we will divide the editors from each year into two groups, those with previous N-SSATS editing experience and those without. We will compare the performance of these two groups on such per case factors as average number of errors found, average editing time, types of edits made, and cost per edited case. To investigate the latter, we will use regression analysis to look at the influence of such variables as highest educational degree, age, previous editing experience, and previous N-SSATS editing experience on editing quality and efficiency.

Improving Response Rates: Consent, Incentives & Caller-ID

Communicating Disclosure Risk in Informed Consent Statements

**Eleanor Singer, ISR, University of Michigan (esinger@isr.umich.edu); Mick P. Couper, ISR,
University of Michigan (mcouper@isr.umich.edu)**

For several years we have been investigating how various ways of communicating disclosure risk and harm to respondents affects their willingness to participate in surveys. These experiments, which used vignettes administered to an online panel as well as a mail survey sent to a national probability sample, have demonstrated that (a) the probability of disclosure alone has no apparent effect on people's willingness to participate in the survey described, (b) the sensitivity of the survey topic has such an effect, and (c) making explicit the possible harms that might result from disclosure also reduces willingness to participate, in both the vignette and the mail experiments.

As a last study in this series we experimented with different ways of talking about disclosure risk in informed consent statements that might more plausibly be used in real surveys, again using vignettes. The study used a 4 (topic) x 6 (confidentiality assurance) design. Two of the topics were sensitive (sex, money) and two were not (work, leisure time). The confidentiality statement assured confidentiality "except as required by law" or "to the fullest extent of the law" or gave an estimated probability of disclosure (one in a million). Three of the statements contained, in addition, the following reassurance: "In our experience at the Survey Research Center, no one, to the best of our knowledge, has ever been

harmful through a breach of confidentiality.” Mode (face-to-face), sponsor (National Institutes of Health), length (15 minutes) and incentive (\$5) were kept constant across the 24 vignettes.

The survey was administered to 7200 members of an online panel by Market Strategies Inc. The presentation analyzes the main and interactive effects of topic sensitivity, type of assurance, and reassurance on the basis of experience on subjects’ willingness to participate in the survey described.

Who’s Cashing In? Demographic Profiles of Incentivized Sample Members

Evan Nielsen, NORC (nielsen-evan@norc.org); Bryan Duff, NORC (duff-bryan@norc.org)

The Survey of Doctorate Recipients (SDR) is a biennial multi-mode panel survey of doctorate recipients in the science, engineering, and health fields from U.S. degree granting institutions sponsored by the National Science Foundation. The 2008 SDR included an early prepaid incentive experiment and a late-stage prepaid incentive offer to nonresponse cases near the end of the field period. The early and late-stage incentive groups are mutually exclusive. The incentivized cases fell into one of four cashing behavior outcomes: 1) incentive cashed, survey complete, 2) incentive not cashed, survey complete, 3) incentive not cashed, survey incomplete, and 4) incentive cashed, survey incomplete. Examining the demographic characteristics of cases in these four outcomes could lead to better forecasting of final incentive costs and identify types of sample members that are particularly difficult to motivate to complete the survey. This paper will present the analysis of the 2008 incentive and survey outcomes for the early and late-stage incentive groups by the following demographic variables: age, gender, race/ethnicity, birth citizenship, doctoral field of study, length of panel tenure, and incentive behavior and survey status in 2006 SDR.

Caller ID Experiments in Telephone Surveys

Eran Ben-Porath, SSRS (ebenporath@ssrs.com); Susan Sherr, SSRS (ssherr@ssrs.com); David Dutwin, Social Science Research Solutions (ddutwin@icrsurvey.com); Melissa Herrmann, ICR/International Communications Research (mherrmann@ssrs.com)

In an effort to maximize survey response, many survey researchers are utilizing customized caller IDs in their projects. This paper explores three specific caller ID experiments. The first was a randomization of two IDs in the Minnesota Health Access Survey, one with the standard ID flag generated by the survey vendor’s phone number, and one with a MN area coded number flagged to the “University of MN.” Outcome measures included overall response rate as well as the rate of answering machines and other dispositions. The second experiment took place within the 2009 Pew National Survey of Latinos. In this experiment we randomized respondents to one of two 1-800 numbers, one of which held the flag of “Pew Survey” with the other as “Estudio de Pew.” Response rates and dispositions were explored by Hispanic density, with the hypothesis that areas of high Hispanic incidence would differ than areas of low Hispanic incidence. The final experiment explored potential differences within a large-city Jewish population demography study. A random subset of respondents received a flag denoting the call as a “Jewish study” with other respondents receiving a non-descript survey vendor ID flag. In addition to participation, differences in self-reported Jewish incidence were explored. Results indicate that, in contrast to expectations, Caller ID does not have a significant impact on participation.

Does One-time Inducements Help Long-Term Survey Participation? - A Case for Survival Analysis.

Kumar Rao, Gallup Inc. (kumar_rao@gallup.com); Allan McCutcheon, University of Nebraska-Lincoln (amccutche@unlserve.unl.edu)

In a recent study involving a mixed-mode experiment to recruit members to a consumer panel, Rao et al. (forthcoming) tested the effect of various response-inducement techniques such as advance letters, monetary incentives, and telephone follow-up on panel recruitment. The experiment was successful in recruiting individuals to a non-incentive driven consumer panel (i.e., members do not receive any form of monetary incentives for their participation in the panel). While the response inducements used were a one-time stimulus to motivate individuals to join the panel, no such response inducements were provided for their subsequent panel participation. After more than a year since the original experiment was conducted, we pose the question: how much time must pass before the effect of this response inducement wears off? An overarching question is how does the propensity to participate in the long run differ across the various treatment and demographic groups from which these individuals were recruited? In this study, we draw upon panel participation data for these individuals to analyze survival rates and the effect of individuals' demographic characteristics on the odds of survival. The study is likely to have important implications for the usage of one-time inducements for long term panel participation.

Citation:

Rao, Kumar, Olena Kaminska, and Allan L. McCutcheon (forthcoming) "Recruiting Probability Samples for a Multi-Mode Research Panel with Internet and Mail Components," *Public Opinion Quarterly* (spring, 2010)

Timing is Everything (Part 3): Examining the Effect of Offering Late-Stage Cash Incentives in a Panel Survey.

Karen Grigorian, NORC (grigorian-karen@norc.uchicago.edu); Deborah Harper, NORC (Harper-Deborah@norc.org); Thomas B. Hoffer, NORC (hoffer-tom@norc.org)

In an effort to maintain response rates, many survey researchers have begun to use monetary incentives to achieve targeted response rates. The Survey of Doctorate Recipients (SDR) is a biennial multi-mode panel survey of doctorate recipients in the science, engineering, and health fields from U.S. degree granting institutions sponsored by the National Science Foundation. The SDR has been trying to determine the best way to optimize limited incentive funds. In the 2003 SDR, a late-stage experiment showed that offering a prepaid incentive not only significantly improved response, but also yielded significantly higher quality data. Building on the 2003 results, the 2006 SDR included a controlled experiment to determine the most effective time to offer a late-stage incentive to different types of sample members. In 2008 SDR, the research team implemented a follow-up controlled experiment to determine the effect of offering an early incentive to two groups, new cohort panel members and those that refused to participate in the 2006 survey, and also offered a late-stage incentive to a randomly selected group of nonresponse panel members at the end of the field period. This paper will compare the response behavior of 24,089 consistently located panel members included in the 2003, 2006 and 2008 survey cycles by late-stage incentive status across the survey cycles to determine if we are conditioning panel members to wait for an incentive offer. We will examine the results by major demographic groups and present an evaluation of unit response, time to response, level of effort (measured by number of contacts) required to obtain a response, and average incentive cost.

International Standards and the Future of Survey Research

Panel: International Standards and the Future of Survey Research.

Reg Baker, Market Strategies International (reg_baker@marketstrategies.com); Tom W. Smith, NORC (smitht@norc.uchicago.edu); Lars Lyberg, Statistics Sweden (lars.lyberg@scb.se); Jim

Wolf, SRC at Indiana University-Purdue University Indianapolis (jamwolf@iupui.edu); Cathy Kneidl, CASRO Institute for Research Quality (Kneidlce@sbcglobal.net)

In April of 2006 the International Organization for Standardization released a global standard covering all of the steps and processes involved in doing market, opinion, and social research. The standard was developed by an international team representing over 20 countries and a mix of organizations including commercial research firms, national standards bodies, and professional associations. The standard's main hallmarks are process standardization and full disclosure of research procedures and outcomes to clients. The standard arose from a widely perceived need for greater transparency in research generally and for a common basis for global research. Since its release over 250 organizations worldwide have certified. While most of those organizations are commercial research firms, Statistics Sweden is now actively pursuing certification. In North America, the Council of American Survey Research Organizations (CASRO) has established a certification body to make it possible for US research organizations in all sectors to become certified. This panel will provide a forum in which to discuss the origins of the standard; its primary features; the benefits of certification; the certification process itself; the likely cost of the effort; and the implications for survey research over the next decade. This session may be of special interest to those contemplating ways for their organizations to participate in AAPOR's Transparency Initiative.

Interviewer Effects & Self-Report Bias

Interviewer Effects on Consenting to Data Linkage on a Longitudinal Survey of a General Population

Emanuela Sala, ISER (esala@essex.ac.uk); Jon Burton, ISER, University of Essex (jburton@essex.ac.uk); Gundi Knies, ISER, University of Essex (gknies@essex.ac.uk)

Linking survey data to administrative data is becoming a common practice in the UK. Data linkage enhances research opportunities as it can provide more data at lower (survey and respondent) costs while, in principle, providing information that could also be used to detect measurement error (i.e., by carrying out validation studies) and therefore improve methods of data collection. In the UK, informed consent must be asked and obtained from respondents in order to link administrative data at the individual level to survey data.

In this paper we explore the interviewers' role in obtaining consent to health data linkage on the 18th wave of British Household Panel Survey. Using a logistic regression model we estimate the effect on consenting of interviewer socio-demographic characteristics, their personality trait ("Big Five"), their job motivation and their attitude to persuading survey respondents (Lehtonen, 1995), controlling for respondent characteristics and interviewer-respondent match. For instance, interviewers who think that they shouldn't try to persuade a reluctant respondent or who are sensitive to possible privacy concerns may be less likely to obtain consent or be less convincing when they try to. Results of this analysis will contribute to clarify issues that are not well researched as well as providing more practical implications for survey designers such as help on briefing interviewers or selecting interviewers who will be asking for consent.

We use a unique matched dataset for this analysis which has never been used before: (i) the BHPS Wave 18 individual interviews, (ii) an interviewer survey which was administered during Wave 18 interviewer briefings and (iii) a unique dataset on all interviewers who ever worked on the BHPS provided by the survey agency.

Representing Seniors in an Online National Probability Panel Survey: Measuring Technology Attitudes and Behaviors

Jordon Peugh, Knowledge Networks (jpeugh@knowledgegenetworks.com); Wendy Mansfield, Knowledge Networks (wmansfield@knowledgegenetworks.com); Tom Wells, Knowledge Networks (twells@knowledgegenetworks.com); Karen Semans, Knowledge Networks (ksemans@knowledgegenetworks.com)

Representation of seniors is often a concern when considering an online panel to conduct research due to their lower access and use of the internet compared to other groups. Using data from KnowledgePanel®, a national probability-based web panel that provides internet access to non-internet households, we will present our analyses on the distinctive technology attitudes and behaviors of seniors compared to other age groups. Moreover, we will also present evidence that seniors provided internet access upon KnowledgePanel recruitment have distinctive technology attitudes and behaviors compared to seniors already having internet access.

Nineteen percent of the adults in KnowledgePanel® are 65 years old and older, comparable to the Census benchmark of 17% for the U.S. population (July 2009 Current Population Survey data). We examined statistical differences across age groups in 15 technology-related attitude statements and identified significant differences between seniors and other age groups on all of these statements. Looking among seniors, we found that those seniors provided internet access upon agreeing to join KnowledgePanel expressed significantly less “pro-technology” sentiment on 12 of the 15 statements compared to those who had prior access. The final paper will present full results which will include an examination of differences across age groups in adoption of personal technologies such as cell phones and online media use.

To measure the impact of panel membership on attitudes and behaviors, we will explore within-person change over time on technology attitudes. KN fielded the same technology attitude items to our panelists beginning in November of 2008 and will again beginning in November of 2009. We expect to have results at both intervals for approximately 4,000 seniors. We will present changes to these measures over time, comparing seniors to younger groups, and seniors provided internet access to those already online prior to joining the panel.

Comparing Proxy to Self-Report Responses of a Civilian Institutionalized Population on a Quality of Life Survey

Alisa M Ainbinder, Mathematica Policy Research (aaainbinder@mathematica-mpr.com) Samuel Simon, Mathematica Policy Research

Numerous well-known household surveys exclude civilian institutionalized populations with disabilities, such as residents of nursing homes, because of the challenges associated with interviewing these populations. As a result, there may be a gap in our understanding of the perspectives of populations who are, according to the American Community Survey figures, significantly older and have a higher rate of disability than the non-institutionalized population. Because this group has a higher incidence of disability than the non-institutionalized population, there is a greater reliance on conducting assisted or proxy surveys. Although there have been some significant studies that compare self-report to proxy responses, little is known about how proxy responses compare for institutionalized populations. Addressing this gap in our knowledge is needed as the research community wants more information about these populations who are frequently excluded from national surveys.

The purpose of this research is to compare proxy and self-report responses to a Quality of Life survey funded by the Centers for Medicare and Medicaid Services and administered to various types of institutionalized populations who are transitioning to community-based care, including the elderly, non-elderly with physical disabilities, people with developmental disabilities, and people with mental illnesses. Approximately 29% of all surveys completed at baseline have been administered with an assisted or proxy interviewee. The first analysis compares assisted, proxy, and participant responses to key survey items and examines differences in number or type of missing data, “don’t knows” or “refusals” of survey items. The second analysis examines whether assisted and proxy responses vary by type of institutionalized population. The results are contextualized within the current literature and provide recommendations for future research on this topic.

Gender of Interviewer: How Does It Influence Responses in a Gender-Focused Survey?

Wendy Wang, Pew Research Center (wwang@pewresearch.org); Jocelyn Kiley, Pew Research Center (jkiley@pewresearch.org); Rich Morin, Pew Research Center (rmorin@pewresearch.org)

Interviewer characteristics such as gender may affect respondents’ answers and produce measurement bias. However, research on when and how gender-of-interviewer affects responses has been limited, and largely inconclusive.

Using data from a 2008 Pew Research Center telephone survey of U.S. adults that focused on gender and leadership, we look at the effect of gender-of-interviewer on responses to various questions about gender roles and characteristics. Preliminary analysis finds significant gender-of-interviewer effects on questions about gender and that these effects are more pronounced among male respondents than female respondents. These findings appear to be consistent with previous research in the area.

Questions analyzed in this paper include: a series of items about the respondent’s preferences for the gender of those in professional roles (e.g., “would you prefer if your surgeon were a man or a woman?”), a battery of questions asking respondents to assess whether some characteristics are more true of men or more true of women (e.g., intelligent, manipulative etc.), and several questions about men and women in political office. Analysis to date suggests that men interviewed by men were more likely than those interviewed by women to prefer male surgeons, doctors, and pilots. They were also more likely to ascribe several positive characteristics to men (e.g., hard-working, intelligent) and negative characteristics to women (e.g., manipulative, stubborn) and to identify men as better at positive traits associated with performance in public office (e.g., standing up for one’s beliefs). This paper will explore these questions as well as gender-neutral questions through multivariate analyses and document the areas where gender-of-interviewer has the most pronounced effect. Additionally, we will explore the interaction of respondent gender and other demographic characteristics with a particular focus on whether these apparent gender-of-interviewer effects differ by age and education of respondents.

How Much of Interviewer Variance is Really Nonresponse Error Variance?

Brady T West, Institute for Social Research (bwest@umich.edu); Kristen Olson, University of Nebraska-Lincoln (kolson5@unl.edu)

Kish’s classical intra-interviewer correlation (ρ_{int}) (Kish, 1962) provides survey researchers with an estimate of the effect of interviewers on variation in measurements of a survey variable of interest. This correlation is an undesirable product of the data collection process that can arise when answers from respondents interviewed by the same interviewer are more similar to each other than answers from other respondents, resulting in correlated response deviations and decreasing the precision of survey estimates. Estimation of this parameter, however, only uses respondent data. Thus, the potential

contribution of variance in nonresponse errors between interviewers to the estimation of rho-int has been largely ignored. Responses within interviewers may appear correlated because the interviewers successfully obtain cooperation from different pools of respondents, not because of systematic response deviations. This study takes a first step in filling this gap in the literature on interviewer effects by presenting a secondary analysis of a unique survey data set, collected using CATI from a record-based sample of divorce cases where official dates of marriage and divorce were known for all sampled cases. The study examines the proportion of between-interviewer variance in key survey measures that is explained by interviewer-level measures of nonresponse error, and shows that significant components of variance due to the interviewer (and thus significant intra-interviewer correlations) are primarily defined by nonresponse error variance across interviewers rather than measurement error variance.

Panel Surveys & Non-Response Issues

Mandatory Reporting: Potential Effects on Retention Rates in a Longitudinal Survey

Kristin Joan Fuller, RTI International (kristinf@rti.org); Melissa Dolan, RTI International (mdolan@rti.org); Keith Smith, RTI International (kesmith@rti.org); Lauren Cohen, RTI International (lcohen@rti.org)

Protection of human subjects, particularly members of at-risk or vulnerable populations, is an essential concern for survey researchers and the Institutional Review Boards that govern them. Little is known, however, about the effects of mandatory reporting of suspected abuse or neglect on retention rates in field data collection studies.

This paper examines the impact of mandatory reporting of suspected child abuse and breach of confidentiality due to child-expressed suicidal intent on retention rates in an at-risk sample. The National Survey of Child and Adolescent Well-being (NSCAW) is a national, longitudinal study of children and families who have come into contact with the child welfare system. Mandated by Congress in 1996, this study is sponsored by the Administration for Children and Families (ACF). Five waves of interview data were collected from the first cohort, NSCAW I. A second cohort was drawn in 2008, with NSCAW II now in its second wave of data collection. Data from both cohorts will be considered in evaluations.

The few studies examining mandatory reporting effects on retention rates suggest limited risk for negative impact on retention. This paper will analyze retention rates of NSCAW respondents in waves following a mandatory report, and whether demographic characteristics (e.g., education, income, race, age of child) or other variables (e.g., type of maltreatment bringing family into the study, child discipline and reports of violence in the household) contribute to the likelihood of continued participation.

Additionally, an exploratory comparison of the number of mandatory reports made during the first wave of NSCAW II (approximately 10 years later) versus reports made in NSCAW I will be conducted in order to examine the impact of increased IRB requirements, and, in turn, potential for effects on respondent retention.

The Unusual Suspects for Survey Non-response: A Follow-up Study with Respondents and Non-Respondents on Their Attitudes and Behaviors for Survey Participation.

Jennie Lai, The Nielsen Company (Jennie.Lai@Nielsen.com); Scott Bell, The Nielsen Company (Scott.Bell@Nielsen.com); Michael Link, The Nielsen Company (Michael.Link@Nielsen.com); Matt Walker, The Nielsen Company (Matt.Walker@Nielsen.com); Leilani Narciso, The Nielsen Company (Leilani.Narciso@Nielsen.com); Teresa Jin, The Nielsen Company (Teresa.Jin@Nielsen.com)

Survey researchers have continued to be challenged with declining response rates coupled with threats of non-response bias due to disproportionate survey participation in recent years. There is an increasing need to curtail this trend by understanding survey participation attitudes for both types of potential participants, respondents and the elusive non-respondents, in order to formulate an effective approach for future surveys. In an effort to gain insight on survey participation attitudes, The Nielsen Company, working on behalf of the Council for Research Excellence, sponsored one of the most comprehensive follow-up studies to investigate whether households that were sampled and cooperated in Nielsen's TV diary measurement differed in meaningful ways from households that did not cooperate. This large-scale study employed a multi-mode methodology using mail, Web, telephone and in-person interviewing during 2007-2008 to achieve the highest response rate possible. The results from this extensive follow-up study uncover the attitudes of these past Nielsen respondents and non-respondents on survey research in general and their perception/experience with the research organization. The first phase of the analysis focused on which attitudinal questions were significant predictors of survey participation. These attitudinal questions focused on some of the usual reasons individuals do not participate in surveys (lack of interest or knowledge on research topic, length of survey, dislike of sharing information, expectation for monetary incentive, etc.). The findings indicated that responses from non-respondents to some of these usual reasons were counterintuitive based on past research findings. The second phase of the analysis is underway to determine if the reasons that are significantly associated with survey participation in the Nielsen sample differ for various household characteristics. This research will explore the attitudinal and demographic characteristics of these households in order to tailor the recruitment strategy to achieve higher response rates and potentially lower the impact of non-response bias.

Locating Efforts and Outcomes in a Longitudinal Study of Older Adults

Jennifer Satorius, NORC (satorius-jennifer@norc.org); Erin Burgess, NORC (burgess-erin@norc.org); Lucie Kalousova, NORC (kalousova-lucie@norc.org); Angela Jaszczak, NORC (jaszczak-angie@norc.org); Laurie Imhof, NORC (imhof-laurie@norc.org)

In order to achieve high response rates and maintain representative samples in longitudinal studies, it is vitally important to be able to locate respondents. The National Social Life, Health, and Aging Project (NSHAP) is a multi-modal, longitudinal study that examines the interaction between health and social relationships in older adults. In 2005 - 2006 during the first wave of data collection, NSHAP completed 3,005 interviews with a nationally representative sample of community-dwelling adults aged 57 to 85. As part of the preparations for the second wave of NSHAP in 2010, NORC began locating efforts early in 2009.

NORC sent a mailing to the Wave 1 NSHAP respondents that included a letter thanking them for their participation in NSHAP and informing them of our intent to return for a follow-up interview in 2010. The mailing also included an NSHAP magnet as a token of our appreciation, and a postage-paid postcard, which allowed the respondent to confirm or update their contact information to send back to NORC. The postcard was designed to seal closed to protect the privacy of the respondent information. The postcard also provided respondents with the option to contact NORC by phone or e-mail in order to update their information. Our current response rate for the postcard mailing is 29%.

This paper will examine the demographic characteristics of responders versus non-responders, mode of response, outcome of response (returned with updates, returned no updates, deceased), as well as those whose mailings were returned undeliverable. Although there is existing literature on locating and tracking respondents in other age groups, there is limited literature on best practices for these activities when

working with an older adult population. We will also discuss geographic mobility specific to an older adult population and in particular how this affects the locating efforts for this population.

An Examination of Panel Conditioning in an Advertising Tracking Study with Cross-Sectional Design

Amanda J Anderson, Fors Marsh Group (aanderson@forsmarshgroup.com); Kara A Marsh, Fors Marsh Group (kmarsh@forsmarshgroup.com); Scott Turner, Fors Marsh Group (sturner@forsmarshgroup.com); Sean M Marsh, Fors Marsh Group (smarsh@forsmarshgroup.com)

Internet panels derived from probability-based sampling are attractive to researchers confronting the limitations of using more traditional modes, such as RDD telephone, to obtain nationally representative estimates (particularly of youth). However, panel conditioning and panel attrition are potential weaknesses that can compromise panel data quality. In this paper we will examine panel conditioning effects in a cross-sectional tracking study on military advertising that involved the reuse of panelists in order to obtain the desired sample sizes.

Data collection for this study began in 2006 and involved quarterly fieldings for two surveys on military advertising with overlapping content. To obtain the desired sample sizes, panel eligibility criteria were negotiated with the vendor such that (a) a panelist could not participate in both surveys during the same quarterly fielding and (b) panelists who participated in one survey would have to wait 9 months before being eligible to participate in that same survey again.

In this paper, data from one of the surveys will be used (n=5584) to examine differences among the following groups: (1) prior participation in same survey (n=1253, 22%); (2) prior participation in similar survey (n=225, 4%); (3) prior participation in same survey AND similar survey (n=999, 18%) and (4) no prior participation (n=3107, 56%).

Data analysis will examine group differences in responses to questions on advertising awareness, brand imagery, and product endorsement. Data quality will also be examined by looking at non-differentiation, a form of survey satisficing, and constructing tests of concurrent validity. The conclusion will offer new insights regarding the boundaries of panel conditioning and provide practical recommendations to researchers who employ Internet-panels for their own efforts.

Reducing Measurement Error Through Cognitive Interviewing

Response Error in Federal Surveys: An Analysis of Q-Bank Summary Data.

Aaron Maitland, National Center for Health Statistics (amaitland@cdc.gov); Heather Ridolfo, National Center for Health Statistics (hridolfo@cdc.gov); Brian W. Ward, University of Maryland (bward@socy.umd.edu)

In an effort to continue advancing questionnaire testing and evaluation, past research has suggested that a database be created that can store and facilitate cumulative knowledge regarding questionnaire pretests. Although this recommendation has been made to include all forms of pretesting (Presser et al. 2004), specific focus has been given to cognitive interviewing (Presser and Wellens 1999), one of the most common methods of questionnaire pretesting (Schaeffer and Presser 2003). With the implementation of Q-Bank (see Miller 2005), there is now a database of cognitive test findings of questions used on federal surveys. With this in mind, the goal of the following study is to use Q-Bank's

summary data to conduct a systematic quantitative analysis of the response error found in those survey questions contained in the Q-Bank database (n=962). Here we ask three research questions. First, what is the prevalence of response error identified by cognitive interviews of questions from federal surveys? Second, what types of response error have been found in federal surveys when using cognitive interviews to pretest survey questions? Finally, how are the characteristics of questions related to response error identified by cognitive interviewing? Descriptive analyses are first conducted to examine both the prevalence of response errors. Bivariate and multivariate analyses are then conducted to examine the effects of various question characteristics on response error, including, the type of information the question elicits (e.g. factual versus subjective), question topic, mode, self/proxy status, and question structure. Finally, the results of these analyses are used to discuss implications for future research.

Should You Take Our Word for It? An Evaluation of Cognitive Pre-test Findings against Other Survey Evaluation Techniques

Alfred D. Tuttle, U.S. Census Bureau (alfred.d.tuttle@census.gov); La Toya Thomas, U.S. Census Bureau (la.toya.thomas@census.gov); Kristin J. Stettler, U.S. Census Bureau (kristin.stettler@census.gov)

The cognitive interview is widely used to evaluate and improve survey questionnaires. Although generally appreciated for its ability to identify potential problems with question comprehension, the cognitive interview method is sometimes criticized for being of questionable validity, owing largely to the fact that the samples used are typically small or not representative. This paper will evaluate the cognitive interview pre-test method as it was applied during the development of the Business R&D and Innovation Survey, a new enterprise-level survey currently being fielded as a two-year full-scale pilot test. We will use evidence generated during the course of routine survey data collection and analysis to assess the findings from cognitive pre-testing during the developmental stage, in terms of detecting potential sources of measurement error. Information such as edit failure rates, item nonresponse rates, imputation rates, telephone follow-up logs, survey analysts' observations about problematic survey items, and findings from debriefings with respondents who have completed the survey will be used to describe the impact of data quality issues for a subset of survey items. These data will be compared with the findings and recommendations from cognitive pre-testing of the survey items in order to attempt to answer the following questions: How well did the cognitive pre-test findings predict observable data quality issues? And what problems became apparent from other analyses that cognitive interviews failed to detect, and why?

Do Respondents Read Those Key Messages in Our Questionnaire Package Cover Letters?: What Comes Out of the Envelope First Matters.

Laurie Schwede, U.S. Census Bureau (Laurel.K.Schwede@census.gov); Anissa Sorokin, U.S. Census Bureau (Anissa.Sorokin@census.gov)

Questionnaire package cover letters often contain critical survey information, but how many respondents read them? If critical messages are not read, the letters are not effectively influencing respondent behavior.

This paper explores the extent to which respondents find and read key messages in questionnaire package cover letters and whether the order in which respondents take materials from the envelope influences that outcome. Data come from three cognitive testing rounds of experimental ACS messages in specially designed cover letters and envelopes aimed at mitigating a potential decline in 2010 ACS response rates during the heightened census environment when some respondents get both the ACS and Census 2010 forms from the Census Bureau.

We unobtrusively observed and recorded whether respondents pulled contents from the ACS envelope with the questionnaire or one of the four other inserts on top and whether they found and appeared to read experimental text in the initial ACS questionnaire package letter. Results were consistent over all rounds and within narrow ranges. Overall, 76% of the respondents opened the envelope with the questionnaire on top and 32%-43% of respondents appeared to find and read the key letter text. We noticed clear differences in who found and read the letter: 18-32% of the questionnaire-on-top group appeared to read or scan it compared to 78% of those with some other insert facing up. The questionnaire-on-top respondents focused on that form and were less likely to go through the other inserts and find and read the letter. The others were more likely to go through the stack and see the letter before the questionnaire. One suggestion is to consider reordering ACS materials in envelopes to make the letter more noticeable. We present results from a Census 2010 cognitive project to see if this pattern may also be found with other mailout packages.

Understanding Measurement Error in the Reporting of Relationship and Marital Status Among Same-Sex Couples

Nancy Bates, U.S. Census Bureau (nancy.a.bates@census.gov); Theresa J DeMaio, U.S. Census Bureau (theresa.j.demaio@census.gov)

In order to accurately portray a population's demographic and social profile, the measures used to produce such profiles must keep up with changes in society and laws. Society and law are changing with respect to same-sex couples. Relationship to the householder is measured the same way in both the Decennial Census and the American Community Survey and is used to define household units which, in turn, are used to establish measures of well-being such as household income and poverty. In 2004, Massachusetts enacted legislation making same-sex marriage legal. Since then, several other states have passed similar laws. Consequently, the number of same-sex couples selecting "husband or wife" is expected to increase as legislation allowing same-sex marriage increases as well.

However, estimates from the Census Bureau's American Community Survey suggest that the number of same-sex couples reporting "husband or wife" is much larger than the actual number of same-sex couples legally married in the U.S. (Williams Institute, 2009; Gates and Steinberger, 2009; O'Connell and Lofquest, 2009). Several hypotheses attempt to explain this phenomenon. One obvious explanation is that some same-sex couples equate their living situation to a marriage regardless of legal status. Alternatively, questionnaire design may play a part – the spouse category is first in the relationship category list on both the Census and ACS forms while unmarried partner is next to last in a long list of fourteen.

This paper reports focus group research to investigate this measurement error. Using participants from same-sex cohabitating relationships, we sought to better understand how same-sex couples answer the current relationship and marital status questions. We also explored what certain terms, definitions, and categories mean to this subpopulation as well as questionnaire design effects. The paper concludes with recommendations of revised questions to be further tested in cognitive interviews and small scale tests.

He Said, She Said: Evaluating Contradictory Cognitive Interview Findings in Two Different Studies of Similar Survey Questions.

Heather Ridolfo, University of Maryland (hridolfo@cdc.gov); Stephanie Willson, National Center for Health Statistics (zex8@cdc.gov)

This paper discusses two different cognitive interviewing reports that arrived at different conclusions when testing the same survey questions. Further, it suggests ways in which readers can evaluate the veracity of competing conclusions, and judge for themselves which to accept. Two reports from the Questionnaire Design Research Laboratory (QDRL) at the National Center for Health Statistics are compared and analyzed. The first report is a study of the Pregnancy Risk Assessment Monitoring System, 2004 (PRAMS). PRAMS is a state-specific, population-based, self-administered survey of women with children between the ages of two and nine months old. Topics center around health issues before, during and after a woman's most recent pregnancy and fifteen cognitive interviews were conducted. The second report is a study of PRAMS 2006 and contains survey items similar to those in 2004. Twenty cognitive interviews were conducted in 2006. Each study was conducted by a different principle investigator at the QDRL, and each arrived at different conclusions after testing similar questions. For example, both the 2004 and 2006 PRAMS surveys had questions asking respondents if they had health insurance prior to becoming pregnant, and if so, what kind (public vs. private). The 2004 report shows no problems or response errors with the insurance questions, while the 2006 report documents a problem related to a faulty assumption about intention to get pregnant. In discussing this and other examples of contradictory findings, this paper shows that the basics of qualitative methodology, including transparency, consistency/coherency, and communicability can be used as criteria for judging the quality of findings in cognitive interview reports.

Survey Sampling: Landline & Cell Phone Undercoverage

Using a Simulation Study to Examine Strategies for Combining Cell and Landline Survey Samples

Tom Duffy, ICF Macro (thomas.duffy@macrointernational.com); Sara Bausch, ICF Macro

(sara.bausch@macrointernational.com); Ronaldo Iachan, ICF Macro

(ronaldo.iachan@macrointernational.com); Bo Lu, The Ohio State University (blu@cph.osu.edu)

Cell phone use has grown quickly in the past decade. According to the National Health Interview Survey (NHIS) estimate, the proportion of adults residing in cell-only households is approaching 20%. Therefore, the representativeness of the traditional RDD landline survey is being questioned due to the increased popularity of cell phone use. Realizing this trend, more and more large state-based surveys have started to include a cell phone supplement sample in addition to an RDD landline sample. Such a design is referred to as a dual-frame survey because samples are drawn independently from two overlapping sampling frames to cover the population of interest. Based on our experience with the 2008 Ohio Family Health Survey (OFHS), we summarize the issues and challenges involved in the design and analysis of dual-frame survey and review the popular estimation methods through a simulation study. The simulation will run several hundred sub-samples of OFHS data to examine the implications of weighting strategies.

Developing Linkage Rules to Reduce Noncoverage

Graham Kalton, Westat (grahamkalton@westat.com)

Noncoverage occurs in nearly all household surveys. It is currently of particular concern for surveys that use the USPS lists as a sampling frame. Linking procedures, such as the half-open interval, have been used since the early years in the development of probability sampling methods to reduce the level of noncoverage. This paper reviews linking procedures that can be used with the USPS lists and in other situations, emphasizing the flexibility of the approach. For area sampling, it will discuss: the half-open interval; a widely-used missed dwelling unit procedure applied with conventional listings of dwellings; a modification that can be applied when the level of noncoverage is higher than usual; and a modification that can be applied when USPS lists are used. With telephone surveys, linking can be achieved through

the use of network sampling. The paper will also describe the use of linkage procedures to “freshen” a longitudinal survey in order to incorporate new entrants to the survey population into the sample at later waves of data collection.

Measuring Undercoverage of the Landline Telephone Population in 1+100-Bank Surveys

Martin Barron, NORC at the University of Chicago (barron-martin@norc.org); Zhen Zhao, Centers for Disease Control and Prevention (zaz0@CDC.GOV)

For many years, survey researchers have assumed that the sampling frame of 1+100-banks for random digit dialing (RDD) landline telephone surveys missed only a small fraction of landline telephone households. In recent years, two studies have results with dramatically different findings: one study reported only 5% of landline households were missed, while the other reported 20% of landline households were missed. To resolve this discrepancy, NORC at the University of Chicago and the Centers for Disease Control and Prevention conducted a study of the 0-bank population using the infrastructure of National Immunization Survey.

We find that estimates of percentage of landline households not covered by 1+100-bank sampling vary substantially depending upon the assumptions one adopts about unresolved lines. We thus detail three different approaches likely to result in an upper and lower bound estimate as well as an approach that we believe is likely to result in the most accurate estimate. In our first approach, we assume that the distribution of unresolved lines is equal to the distribution of resolved lines. In our second approach, we treat the distribution of unresolved lines as a two-step process where working status is estimated separately from residential status. In our third approach, we mirror Approach One but assume that unresolved refusals or hang-ups are residential phone numbers. Our estimates of the undercoverage of landline telephone households from these approaches range from 6.7% (Approach One) to 15.6% (Approach Three). We believe the rate is likely closer to 8.1% (Approach Two). We discuss the implications of this rate of undercoverage for 1+100-bank RDD surveys.

Assessing Cell Phone Noncoverage Bias Across Different Topics and Subgroups

Leah Melani Christian, Pew Research Center for the People & the Press

(Ichristian@pewresearch.org); Scott Keeter, Pew Research Center (skeeter@pewresearch.org);

Kristen Purcell, Pew Internet & American Life Project (kpurcell@pewinternet.org); Aaron Smith, Pew Internet & American Life Project (asmith@pewresearch.org)

Although many national telephone surveys are now conducted using dual frame samples that include cell phones, many other surveys – including most state and local surveys - still do not interview people on cell phones because of cost considerations or difficulties with sampling. As the number of households without landline phones continues to grow, the potential for noncoverage bias also gets larger. Only a few examples of a significant bias have appeared in the literature, but previous research indicates that the potential for bias is greater for certain topics and population subgroups than for others. For example, noncoverage bias has been identified for certain health behaviors among young adults, for presidential vote preference among adults ages 30-39 and – perhaps not surprisingly – for questions about cell phone usage.

We update and extend a previous Pew Research Center study of noncoverage bias by analyzing questions included in a variety of surveys conducted in 2008 and 2009. To assess the size of the bias, we compare weighted estimates from landline respondents to those obtained from combined samples of landline and cell respondents, using appropriate statistical tests for overlapping samples. The surveys cover a wide range of topics including public policy issues, personal and national economic ratings,

foreign policy views, political values, attitudes about technology, Internet usage and communications behaviors, religious and social values, and attitudes about science.

Although little evidence has emerged to indicate that “cell mostly” respondents are underrepresented in landline surveys, we also will evaluate whether telephone usage patterns among dual households (with both landlines and cell phones) are associated with substantive responses.

How Much Coverage Does an RDD Frame Really Provide?

Robin Gentry, Arbitron (robin.gentry@arbitron.com); Alan Tupek, Arbitron, Inc. (alan.tupek@arbitron.com)

Arbitron, a provider of media ratings, currently uses a hybrid sampling approach to cover the population in the metro areas it surveys. An RDD landline sampling frame is used to cover households with one or more landline telephones, and an address frame is used to identify and cover cell phone only (CPO) households. While including CPO households has helped Arbitron improve its coverage of the population, especially persons 18-34, recent research studies by companies that provide the RDD landline sampling frame have raised questions about the coverage of households with landline telephones. With the expanding phone service options for consumers, including cable companies and voice over internet protocol (VOIP), the traditional RDD landline sample frame has come under increasing scrutiny among survey organizations. Estimates of the number of residential phone numbers that may be outside the 1+ 100-banks frame have ranged from 5 to 20 percent (see www.surveypractice.org/ September 2008, January 2009)

Arbitron asked two RDD sample providers to match phone numbers obtained from the address frame study to the traditional 1+ RDD sample frame. In Summer 2009, roughly 4000 numbers not matching the 1+ RDD frame were re-contacted to determine their phone number type (cell or landline) and additional information on the company providing their phone service, cable and satellite service, and household demographics. This presentation will provide results from this test and will show how the data collected, in combination with other sources of coverage data can be used to estimate levels of undercoverage in the RDD frame.

AAPOR Concurrent Session B

Sunday, May 16, 2010

10:15 a.m. - 11:45 a.m.

Behavior Coding & Cognitive Interviewing

Cognitive Interviews in Establishment Surveys: Getting Quality Feedback in a Diverse Setting of U.S. Ferry Operators

Kenneth Steve, Bureau of Transportation Statistics (Kenneth.Steve@dot.gov); Joy Sharp, Bureau of Transportation Statistics (Joy.Sharp@dot.gov)

As mandated by legislation, the Bureau of Transportation Statistics (BTS) conducted its first National Census of Ferry Operators (NCFO) in 2006 and subsequently conducted a biennial update in 2008. Data collected from the approximate 260 ferry operators currently operating in the United States were used to establish and maintain a national ferry database containing information regarding routes, vessels, passengers and vehicles carried, funding sources, etc. As with many surveys or censuses of businesses, experience from the last two iterations of the NCFO has demonstrated a wide disparity in the size, operations, and recordkeeping practices of ferry operators. In addition, review of the questionnaires and resultant data has indicated numerous data quality problems most likely resulting from unclear questions, inconsistent information requests, and other differences resulting from using one questionnaire to collect information from a very diverse population of operators.

In preparing for the 2010 NCFO, BTS has implemented a quality improvement plan that will include a series of cognitive interviews to evaluate the effectiveness of the current census instrument and explore areas of difficulty in the cognitive response process. These cognitive interviews will be conducted in February and March of 2010 and will take place on-site at the ferry operator locations, encompassing multiple geographic locations and sizes/types of operations. Expanding on existing cognitive research methods for establishment surveys, this paper will thoroughly document the revisions to standard cognitive interviewing methods, describe the research application, and summarize the results.

Utilization of Cognitive Interviewing For Development of Family Bonds Survey Questionnaire

Jiri Remr, Charles University in Prague, Faculty for Social Sciences (remr@centrum.cz)

Paper presents how the technique of cognitive interviewing could be utilized for finding proper indicators and for detection of usual problems with question format and question wording (i.e. comprehension, matching respondent's role perception with the reality, response process and social desirability response bias). This effort is presented within the context of family bonds research.

Research task of our team was to re-establish the theory of family bonds originally prepared by Bengtson within the context of American society. He distinguishes six different dimensions of the family bond using a simple set of approx. 4 – 5 indicators describing each dimension. It is always a difficult task to transfer the theoretical model developed in one cultural context into different one. And if it is the case, then special attention has to be paid to real meaning of each indicator proposed to measure certain type of attitude. Focus groups already conducted by our team proved that the six-dimensional concept perfectly fits with

the mind-map of the respondents – representatives of Central-European families. Even though the general picture could remain the same, actual indicators had to be replaced because of low perception of their relevance and/or because of incomprehension.

Proposed paper opens the methodological options that were considered for application during preparatory stage of the research inquiry; our experiences with cognitive laboratory interviewing in contrast with field-pretesting are differentiated at this point. Paper then describes in detail the approach taken and specifies actual research design; special attention is paid to importance of combination of qualitative and quantitative approaches. Finally, the paper shows how proper indicators were found utilizing the cognitive interviewing technique. Summary of our empirical evidence together with major hints is a luscious part of the paper.

Using Verbal and Paralinguistic Behaviors to Identify Mapping Difficulties in Responses to Self-Reported Health Questions

Dana Garbarski, University of Wisconsin-Madison (dmgarbarski@wisc.edu); Nora Cate Schaeffer, University of Wisconsin-Madison (schaeffe@ssc.wisc.edu); Jennifer Dykema, University of Wisconsin Survey Center (dykema@ssc.wisc.edu)

The self-reported health question (e.g. “How would you rate your health: excellent, very good, good, fair or poor?”) summarizes a wide range of information on health status, and is widely used to measure health because it predicts morbidity and mortality well (Idler and Benyamini, 1997). The social-information-processing model of the response process suggests mapping difficulties as one source of problems for respondents, and these problems may manifest themselves in the interaction between interviewers and respondents. We use data on respondents’ illness and physical functioning to identify respondents who are likely to find it difficult to map their answer onto the self-reported health response options. Interactional behaviors that indicate mapping difficulties are more likely to occur when respondents have inconsistent combinations of illness and functioning such as high illness and low functioning versus low illness and high functioning.

We examine features of interviewer-respondent interaction that may be associated with response error, task difficulty, and cognitive ability (Draisma and Dijkstra, 2004; Schaeffer et al., 2008). Measures include respondent response latencies, expressions of uncertainty and qualifying answers, and interviewer follow-ups. These interactional behaviors can also be viewed as by-products of the information processing that occurs when a respondent answers a survey item (Fowler and Cannell, 1996; Holbrook, Cho, and Johnson, 2006).

Ultimately, we seek to explain in part the variation in responses to the self-reported health question by demonstrating that the features of the interaction that produced the answer reveal important information about respondents’ health status not captured by the self-reported health question.

We analyze interviewer-respondent interaction of the self-reported health question from 360 digitally recorded interviews of older adults in the Wisconsin Longitudinal Study. We coded interaction in Sequence Viewer using an elaborate scheme that permits analysis of very specific respondent and interviewer interactional behaviors such as response latencies, disfluencies, and probing.

Early Responders, Late Responders & Non-Responders

Survey Non-Response Among American Community Survey’s Chinese Respondents.

Anna Y. Chan, U.S. Census Bureau (anna.y.chan@census.gov); Yuling Pan, U.S. Census Bureau (yuling.pan@census.gov); Aref Dajani, U.S. Census Bureau (aref.dajani@census.gov)

Prior qualitative research using data from cognitive interviews revealed that, as a group, Chinese participants are more survey illiterate and are more likely to provide indirect and ambiguous responses to the survey questions when compared to their English speaking counterparts (Chan and Pan 2009, Chan and Pan forthcoming, Pan 2005, Pan 2008.) It is important to understand whether such issues observed during qualitative studies would translate to an overall higher level of survey item non-response among non-English speaking Chinese respondents than it would among their English speaking counterparts. The major goal of this study is to provide empirical evidence to the survey response behaviors that have been observed during cognitive testing studies. In this paper, we examine whether the survey item nonresponse rates among Chinese participants of the American Community Survey (ACS) are significantly higher than those of their English-speaking counterparts. The 2005, 2006, and 2007 ACS data will be used for the analyses of this study. Currently, the ACS questionnaire is provided in English only. The identification of potentially problematic survey items, defined as survey questions with significantly higher item non-response rates for Chinese respondents, will inform future planned cognitive testing studies to evaluate language guides developed to help non-English-speakers to fill out the ACS questionnaire. Findings will shed light on questionnaire design and evaluation for multilingual surveys.

More than Fashionably Late: An Analysis of the Late Responders in the 2008 Survey of Doctorate Recipients

Shana M Brown, National Opinion Research Center (brown-shana@norc.org); Lance Selfa, National Opinion Research Center (selfa-lance@norc.org)

The National Science Foundation-sponsored Survey of Doctorate Recipients (SDR) is a biennial multi-mode panel survey of doctorate recipients in the science, engineering, and health fields from U.S. degree-granting institutions. Like many panel surveys, the SDR attempts to keep within a particular data collection time period while still reaching a federally-specified response rate. For the SDR, the goal is to obtain 80 percent response rate within an approximately 9-10 month data collection period. In the 2008 SDR, the final 19 weeks of a 10 month data collection period were spent obtaining the final 15 percent, or about 4,500 cases (out of approximately 30,000 final completes). Who were these respondents who received the same mailings, telephone calls, and emails as everyone else, but were still late to the party? While the SDR has looked at sampling characteristics of late responders in previous research, here we will examine characteristics based on survey responses in the 2006 survey, such as number of children, marital status, salary, supervisory roles, occupation, hours worked, and so on. Thus this analysis will focus on 2008 late responders who had also participated in the 2006 SDR. Perhaps we can see if the late responders are simply busy with their home and work lives as opposed to being reluctant to participate. We will identify late responders based on their time to complete, taking into account time spent to locate them. We will analyze the characteristics of these respondents compared to “early” responders to determine what, if anything, can be done to get these late responders to complete the 2010 survey in a more timely fashion.

Differences in Early and Late Responders: Findings from a Military Web-Based Community Survey. Jyothsna Prabhakaran, ICF International (JPrabhakaran@icfi.com); Christopher Spera, ICF International (CSpera@icfi.com); Laura M Leach, ICF International (laura@leaches.net); Rachel Foster, U.S. Air Force (rachel.foster@LACKLAND.AF.MIL)

Several recent studies have investigated the relationship between survey response time (i.e., early or late response), respondent demographic characteristics and their answers to key survey questions, as well as

the impact of survey reminder follow-up. These studies have shown that early and late responders often differ in their demographic characteristics as well as in their responses to key attitudinal questions. Often the highest response rate increase follows the initial invitation and various reminder methods, including reminder emails or follow-up mailings which are used to increase response rates throughout the duration of the fielding period. It is important to understand the impact of gathering additional "late" responses through such follow-up campaigns on any potential nonresponse bias. This study examines the results of an anonymous web-based survey, deployed by ICF International, of U.S. Air Force (USAF) Active Duty and civilian spouses at 80 USAF bases worldwide. During the eight-week fielding period, Active Duty members received survey invitations and eight reminders sent via email. Spouses received a 8.5" X 5.5" invitation postcard followed by two 6" X 4" postcard reminders. A total of 56,135 Active Duty members responded, 78% within the first four weeks and 22% in the last four weeks, and a total of 12,342 civilian spouses responded, 64% within the first four weeks, and 36% in the last four weeks. For Active Duty, the average late responder was more likely to be of lower enlisted grade (E1-E4) and female, while no differences were found between demographics among spouses. Both Active Duty and spouse late responders were more likely to rate community and personal satisfaction more positively than early responders. These results suggest that early and late responders may differ, and such potential differences should be considered when designing survey administration fielding and incorporating reminders to encourage the non-responder population to participate.

The National 2009 H1N1 Flu Survey: Rapid Data Collection and Early Responder Analysis

James A. Singleton, Gary L. Euler

National Center for Immunization and Respiratory Diseases

Centers for Disease Control and Prevention

Kirk Wolter, Kennon Copeland, Nicholas Davis, Nada Ganesh

NORC at the University of Chicago. Kirk M Wolter, NORC at the University of Chicago (wolter-kirk@norc.org); Nicholas D Davis, NORC at the University of Chicago (davis-nicholas@norc.org)

The National 2009 H1N1 Flu Survey is a nationwide, random digit dial (RDD) survey conducted by NORC at the University of Chicago on behalf of the Centers for Disease Control and Prevention, with the goal of monitoring vaccination rates among children and adults for the 2009 H1N1 influenza virus (sometimes referred to as "swine flu") as well as seasonal influenza. Because of the rapidly evolving influenza pandemic and the need for information to enable nationwide management of vaccination resources, the survey was designed to collect and deliver data on a weekly basis. Since survey cases must be completed relatively soon after being sampled with a relatively short data-collection period, the data are potentially subject to a non-response bias (i.e., a bias arising from the use of data collected primarily from early responders).

NORC developed a procedure for collecting and processing data as quickly as possible while minimizing the risk of early responder bias. This paper describes the sampling options that were considered and the merits of the chosen sampling scheme with regard to the minimization of early responder bias. We describe a method for the weighting of survey cases that aims to achieve a stable balance of time-in-sample effects across survey weeks.

We compare estimated vaccination rates and other key statistics across multiple panels of data and test for significant differences between the panels. Using logistic regression models, we quantify the effect of time-in-sample on the probability of vaccination. The analysis reveals no substantial early responder bias in the survey estimates obtained from the sampling design and weighting procedure, which in turn suggests the potential utility of the design for other surveys that require a rapid response.

Exploring the Diversity of Public Opinion

Attitudes Toward NAFTA in Continental Perspective

Timothy B. Gravelle, Gallup (tim_gravelle@gallup.com)

The United States and Canada have been formally committed to free trade for more than 20 years, with most of this time being under the trade regime established by the North American Free Trade Agreement (NAFTA). Public support for NAFTA in both countries and views of its effects have, however, remained mixed, and have also varied over time. Drawing on 2008 and 2009 data from the Gallup World Poll, this paper endeavors to explain why Americans and Canadians hold positive (or negative) views of NAFTA and its effect on the national economy. It evaluates several hypotheses relating to the effects of opinions about the national economic climate, one's personal economic situation, impressions of the other country and demographic characteristics on impressions of NAFTA.

Drafting a Response: Demographic and Item Context Effects on Support for Mandatory National Service Options

William Bryan Higgins, ICF International (bhiggins@icfi.com); Randall K. Thomas, ICF International (randall.k.thomas@gmail.com); John F. Kunz, ICF International (jkunz@icfi.com)

Since 2003, bills have been periodically introduced into the U.S. Congress to reinstitute the draft to support increased military preparedness due to the stresses placed on the all-volunteer military active in two distinct war zones. Though the last draft ended in 1973 and was for men only, many are calling for women to be equally subject to national service since they have been ably serving in combat and combat support units. While some proponents of the draft have indicated the need is for military service only, others have thought more broadly of a required national service, of which military service may be one option. We conducted a web-based survey to study the extent of public support for mandatory national service. Respondents were randomly assigned 1 of 5 mandatory service scenarios for 18 year olds: 1 year military only; 1 year military or national service; 2 year military only; 2 year military/national service; or 1 year military/2 year national service. Respondents indicated how much they would support or oppose the service option for males and females separately. The scenarios with a national service option had much more support than military-only options. In addition, support for mandatory service for males averaged 5% higher across all options. We will also report on differences in support for mandatory service due to region, sex, age, education, and party ID.

Understanding How Consumers Cope with Food Safety Concerns: Using Qualitative Panel Data to Obtain a Complete Picture

Julia R Pennington, University of Wisconsin-Eau Claire (penninjr@uwec.edu); Kumar Rao, Gallup Inc. (Kumar_Rao@gallup.com); Michael Krush, North Dakota State University (michael.krush@ndsu.edu); Dwayne Ball, University of Nebraska-Lincoln (dball1@unl.edu)

American consumers are increasingly concerned about the microbial safety of the fruits and vegetables they consume. A random survey of 3619 consumers from the Gallup Panel performed in October 2008 investigated the extent to which American consumers were concerned about the microbial safety of produce and how they coped with those concerns in the food marketing channel. Quantitative analysis found that consumers who were moderately to moderately highly concerned about microbial safety used a variety of coping methods. However, it was unclear if the individual coping methods, as expanded on by consumers through several qualitative open-ended questions, were correlated with other individual

consumer practices and demographics. To investigate the potential relationship, the survey's qualitative information was analyzed using qualitative software with three independent coders. The inter-rater reliability between the coders was calculated. The qualitative results were then correlated with the quantitative results to obtain a fuller picture of how consumers cope with concerns in the food marketing channel. This paper discusses the findings from this study and the implications of our findings for marketing theory, marketing practice, and further research.

Health Insurance and Place: A Multilevel Analysis of Health Insurance Coverage in Texas.
Kathy Krey, Baylor University (Kathy_Krey@baylor.edu); Jodien Matos, Baylor University (Jodien_Johnson@baylor.edu)

The health of any nation is inherently connected to its health care system. While reform may be underway, the American public currently relies on health insurance as their gateway to care. However, insurance has become increasingly expensive leaving some 45 million Americans without adequate coverage, including 7 million children. Insurance coverage is not uniform across geographical locations with Texas ranking highest in the nation for percent uninsured. Not only is 24% of the population in Texas uninsured, but Texas is a rapidly growing state with a high proportion of Hispanics, immigrants, and self-employed, all factors noted to influence insurance coverage. This research uses the American Community Survey and presents a closer look the uninsured within the state of Texas, where there are a number of confounding factors known to influence insurance attainment, to further explain the relationships between insurance and community characteristics. Using Hierarchical Linear Modeling we find that individual predictors of insurance attainment are mediated at the community level.

Factors Related to Trust in the Security of Electronic Medical Records (EMRs).
Sarah Cho, Kaiser Family Foundation (SarahC@kff.org); Sasha Buscho, Kaiser Family Foundation (SashaB@kff.org)

Although the public overwhelmingly agrees in the importance of having an EMR system, they remain skeptical that their medical records would remain confidential if stored electronically. A March 2009 poll conducted by researchers at National Public Radio, Harvard School of Public Health, and the Kaiser Family Foundation found that a majority of the public (59%) reported they were not confident their personal health information would be secure in an EMR system. Data from this national survey of 1,238 respondents was utilized to conduct logistic regression analysis of the demographic characteristics, opinions and health care experiences related to trust in the privacy of EMRs. The results indicate that those without a high school diploma and the uninsured are less likely to report confidence in EMR security, while those with first-hand experience accessing their medical records online are more likely to report trusting EMRs to protect their confidential information. In addition, those who report positive views about the potential effect of EMRs on health care cost, quality and coordination are more likely to report trusting the confidentiality of EMRs. These findings have implications for policymakers working to garner public support for EMR implementation.

Interviewer Characteristics & Maintaining Data Quality

Using Computer Audio Recorded Interviewing (CARI) to Detect and Reduce Total Survey Error.
Tom W. Smith, NORC (smitht@norc.uchicago.edu)

CARI was used on the 2008 General Social Survey. Its use is examined from the following perspectives: 1) as a means for validating interviews, 2) to detect interviewing errors and develop techniques to

improve interviewing, 3) to improve the coding of open-end questions such as on occupation and industry, and 4) to carry out linguistic and discuss analysis of both interviewers and respondents.

Is Success on the Doorstep Correlated With the Magnitude of the Interviewer Design Effect?

Patrick Sturgis, University of Southampton (p.sturgis@soton.ac.uk); Ian Brunton-Smith, University of Surrey (i.r.brunton-smith@surrey.ac.uk); Joel Williams, BMRB Limited (Joel.Williams@tns-bmrb.co.uk)

Evidence is now beginning to accumulate which shows that various aspects of interviewer attitudes, personality, and behaviour are predictive of success in achieving contact and cooperation on the doorstep. For instance, openness, optimism, trust and cognitive flexibility are all dispositions which are thought, at least in part, to underpin successful doorstep approaches. Yet it has long been recognized that the characteristics which might advantage an interviewer in obtaining contact and cooperation might, for similar reasons, also lead to the elicitation of more variable responses during the interview. Thus, while recruiting certain types of interviewers and training them to adopt particular strategies might serve to reduce bias through raising response rates among hard-to-reach and reluctant sample members, total survey error might be unaffected (or increased) due to their effect on the variance of survey estimates. Drawing on face-to-face interview data from three years of the British Crime Survey, a large multi-stage probability sample of the British population, this paper uses cross classified multilevel models with a complex error structure to examine how the size of the interviewer design effect varies as a function of historical measures of interviewer success in obtaining contact and cooperation on the doorstep. This is assessed across a range of different question types, and conditions on the potentially confounding influence of area clustering to provide a detailed account of the extent that different types of interviewers are associated with higher or lower levels of error.

Non-verbal Correlates of Rapport in Face-to-Face Survey Interviews: An Analysis of Interviewer Behavior

Brooke Foucault Welles, Northwestern University (b-foucault@northwestern.edu)

Because rapport may lead to increased coordination or cooperation between interviewers and respondents, it is widely considered to be important for survey interviews. However, previous empirical studies of the effect of rapport on response behavior have been inconclusive. This may be because there is little consensus among survey researchers about how to define or operationalize survey rapport. Further, the survey literature lacks an investigation of how an interviewer's non-verbal behavior affects a respondent's perception of rapport, despite considerable evidence from outside the survey community that rapport is primarily established non-verbally.

To fill that gap, this study examines the relationship between interviewers' nonverbal behaviors and respondents' evaluations of survey rapport. To do so, we videotaped four professional interviewers administering face-to-face surveys to two respondents each (total interviews = 8). In a post-survey questionnaire completed by each respondent, five interviews were rated higher-rapport and three were rated lower-rapport. All of the interviews were coded for three interviewer non-verbal rapport behaviors: smiling, nodding, and direct gaze. A nominal contingency table analysis revealed a significant positive relationship between interviewer smiling and respondent-assessed rapport ($\phi = 0.112$, $p < 0.001$) and interviewer nodding and respondent-assessed rapport ($\phi = 0.091$, $p < 0.001$). In addition, a significant negative relationship between direct gaze (where the interviewer looks directly at the respondent's face) and respondent-assessed rapport ($\phi = -0.077$, $p < 0.001$) was revealed. These results suggest that higher-rapport interviewers smile and nod more frequently than lower-rapport interviewers, and higher-rapport interviewers look at their respondents less frequently than lower-rapport interviewers. Implications

for interviewer training, virtual interviewer design, and directions for future research on the relationship between rapport and response validity are discussed.

Interviewer Characteristics Associated With Productivity, Cost Efficiency, and Retention.

Susan Mitchell, RTI International (smitchell@rti.org); Wandy Stephenson, RTI International (wstephenson@rti.org); Kristine Fahrney, RTI International (fahrney@rti.org); Matthew Strobl, RTI International (mstrobl@rti.org); Barbara Bibb, RTI International (bibb@rti.org)

The costs associated with recruiting and training field interviewers for in-person surveys are significant, often amounting to thousands of dollars per interviewer depending on the location, mode of delivery, and duration of the training. For this reason, it is desirable to hire and train interviewers who will remain on the study through its completion and perform cost efficiently in the field. In low-income communities, interviewer performance and retention can be especially challenging because of concerns for personal safety, the frequent sensitivity of the research topics, and the high mobility of the population. Hiring interviewers with characteristics associated with high productivity, cost efficiency, and retention helps to minimize survey costs through reduced interviewer attrition and lower labor costs per interview.

This paper examines interviewer characteristics that may be associated with survey costs in low-income communities. These characteristics include age, gender, previous experience as a field interviewer on this or other studies, performance during training, rate of pay, and whether interviewers live in the area in which they work. We address the following research questions: (1) is paying higher rates of pay associated with higher productivity and cost efficiency in the field, (2) do interviewers who live in the area where they work have higher productivity and cost efficiency, (3) do interviewers with prior interviewing experience have higher productivity and cost efficiency, and (4) do interviewers who perform better during training have higher productivity and cost efficiency. In addition, we profile interviewers who left the study to identify characteristics that may be predictive of attrition. Findings are based on the Study of Community Family Life, and in-person survey of about 4,000 households in six low-income urban areas, sponsored by the Administration for Children and Families, Department of Health and Human Services.

Economic Recession: Effects on Telephone Survey Staffing and Productivity

McKinlay Jeannis, RTI International (mjeannis@rti.org); Tamara L Terry, RTI International (tterry@rti.org); Laura Flicker, RTI International (lflicker@rti.org)

The downturn in the economy that began in 2007 has had a profound impact on people across the country. Joblessness has risen in all states and resulted in a large pool of skilled unemployed persons. The recession has been difficult for most industries; however the impact on telephone survey research may in fact be positive. The economic downturn has made it easier to recruit staff with a higher level of education and reduce attrition.

To explore the impact that the economy may have on telephone data collection, we assessed two studies within RTI's Call Center Services (CCS) before and during the economic recession of 2007-2009. In order to examine how hiring and retaining skilled staff has affected survey operations in our call center, we correlated the number of calls to complete with interviewer characteristics, such as gender and education. Preliminary results indicate that in the past two years, RTI's CCS interviewers have higher levels of education and we have retained them for longer periods of time.

Non-Response Issues: Interviewer Effect & Other Participation Factors

Preferences for Interviewer Dialect Use and Race Among African American Health Survey Respondents

Rachel E. Davis, University of Michigan (reda@umich.edu); Jennifer G. Nguyen, University of Michigan, Dept. of Linguistics (jgnguyen@umich.edu); Jennifer J. Hawkins, University of Michigan, Internal Medicine-Nephrology (jenjoyce@med.umich.edu); Karen E. Peterson, University of Michigan (karenep@umich.edu); Sara W Schlotterbeck, University of Michigan (sarasch@umich.edu); Ken Resnicow, University of Michigan (kresnic@umich.edu)

This presentation will share findings from a qualitative study exploring African American health survey respondents' attitudes about African American interviewers' use of two dialects, American Standard English (ASE) and African American English (AAE). ASE is the most used English dialect in the U.S., but many African Americans use AAE and may prefer its use as an expression of in-group identity and pride. Researchers have noted potential for miscommunications between ASE and AAE speakers that may influence the survey interview. This study investigated how differences in language use may influence health survey respondents' participation in health surveys, cognitions when answering survey questions, estimation of socially desirable responses, communication of answers, and comfort and trust in interviewers and the treatment of survey data. Respondents were also queried about their preferences for African American interviewers. In-depth, qualitative interviews were conducted with 28 adults representing a range of dialect use from Detroit, MI, and Toledo, OH, between May-August 2009. Pre-recorded samples of interviewer speech demonstrating varying levels of ASE and AAE were used to stimulate discussion about language use during different types of interviewing tasks: recruitment, informed consent, and administering sensitive and non-sensitive survey items. The interviews were digitally recorded to permit linguistic coding of respondents' dialect use and to facilitate analyses using qualitative data analysis software. Early analyses indicate complex interactions between respondents' own dialect use and their attitudes about and preferences for dialect use by African American interviewers during a health survey. In addition, several processes have been identified, such as high language self-monitoring and a tendency for some respondents to translate an interviewer's utterances into AAE in their heads and translate their responses back to ASE before vocalizing them. Major themes and findings from this research will be presented, with possible implications for the dialectical skill sets of health survey interviewers.

To Agree or Not to Agree: Effects of Spoken Language on Survey Participation Decisions

Frederick G. Conrad, University of Michigan (fconrad@isr.umich.edu); Jessica Broome, University of Michigan (jsbroome@umich.edu); Jose Benki, Michigan State University (benki@msu.edu); Robert Groves, U.S. Census Bureau (surveygroves@gmail.com); Frauke Kreuter, University of Maryland (fkreuter@survey.umd.edu); David Vannette, University of Michigan (daleva@umich.edu)

When potential respondents consider whether or not to participate in a telephone interview, they have very little information about the interviewer, aside from how he/she sounds, speaks and interacts. Yet interviewers vary widely in how often their invitations lead to participation, suggesting that potential respondents may give considerable weight to interviewers' verbal attributes. We report a study that investigates the relationships between interviewer-answerer spoken interactions, interviewers' historical response rates, and the outcome of specific telephone survey invitations: agree-to-participate, scheduled-callback, refusal and hang-up. We explore these relationships in a corpus of approximately 1200 audio-recorded contacts, sampled from five studies conducted by the University of Michigan Survey Research Center for which each utterance is transcribed, acoustically measured and coded in detail. We examine the impact of a variety of conversational events such as disfluencies, pauses, interruptions, interviewer apologies, procedural questions by the answerer, speech rate, and pitch on success of each invitation. So far it appears that perfectly fluent interviewer speech is not an asset; interviewers produce more ums and

uhs, i.e. are less fluent, when they are more successful, i.e., produce more agrees and scheduled-callbacks than refusals and hang-ups. In addition, interviewers interrupt the answerer more when they schedule a callback, perhaps reflecting their attempt to salvage the contact, but answerers interrupt interviewers most often when they ultimately refuse to participate, presumably attempting to end the call. When answerers agree to participate, they display more “backchannels,” i.e., utterances such as uh huh or I see which are generally assumed to be a signal to the speaker that the listener is still listening. We conclude by discussing next analytic steps (e.g., including covariates such as the study, the interviewer and the specific householder in our models), as well as practical implications for interviewer hiring and training.

Interviewer Personality and Survey Error: Interviewers' Conscientiousness as Predictor for Item- and Unit Nonresponse in a Panel Household Study

Michael Weinhardt, German Institute for Economic Research, Socio-Economic Panel Study (mweinhardt@diw.de); Frauke Kreuter, University of Maryland, Joint Program in Survey Methodology (fkreuter@umd.edu); Jürgen Schupp, Freie Universität Berlin, German Institute for Economic Research (jschupp@diw.de)

In household panel surveys, interviewers' first task is to secure respondents' cooperation, i.e. initial response and response over time. A second task is to collect and record precise information on the respondents. Previous studies have shown that the 'Big-Five' personality trait conscientiousness is significantly related to general indicators of job performance. In this paper we look at whether this holds for survey interviewers also. The data used to investigate this problem is unique in the sense that it did not have to rely on data provided by fieldwork agencies, but information on interviewers that was collected during a separate study. In December 2006, a survey of all current interviewers of the German Socio-economic Panel Study (SOEP) was conducted and 546 interviewers responded to a 10-page paper questionnaire. This included a short version of the Big Five Inventory, a scale to assess the 'Big Five' personality traits including conscientiousness. Also in 2006, a refreshment sample for the SOEP was drawn and respondents interviewed for the first time, giving the opportunity to study initial unit response. With this data, we examined the effects of interviewers' conscientiousness on initial response, household attrition and item nonresponse by linking survey data from the interviewers with household and individual level information on respondents. Using a multilevel regression framework, we find that interviewer conscientiousness is not related to initial response in 2006, but is significantly related to overall item nonresponse in 2006 and to household attrition in 2007. Results will be discussed in terms of their usefulness for the recruitment and training of interviewers, highlighting the importance of conscientiousness in an interviewer's approach to work.

High Dimension Multiple Imputation: Missing Blood Glucose Values in the Epidemiology of Diabetes Interventions and Complications Study

Michael D Larsen, George Washington University (mlarsen@bsc.gwu.edu); John M Lachin, George Washington University (jml@bsc.gwu.edu)

Increasingly in large scale surveys it is possible to collect genetic information on subjects. The need to address high dimensional imputation is of increasing concern in survey research. The Epidemiology of Diabetes Interventions and Complications (EDIC) study is following subjects from the Diabetes Control and Complications Trial, which compared conventional (n=729) and intensive (n=711) treatment for type 1 or insulin-dependent diabetes. Reported findings have had fundamental impacts on treatment in practice. Recent work has found genetic associations to health measurements and diabetes-related conditions. Missing data in blood glucose profile measurements (7 measurements in a day, one day per quarter over nine years) compromises the ability to evaluate the association of distributional

characteristics, e.g., variation, of glycemia with long term outcomes. Multiple imputation (MI) using chained equations is a feasible and realistic approach to filling in the missing values and enabling certain analyses. To implement the MI strategy, one must use appropriate models for imputation. The number of glucose values (more than 250) and the high fraction missing (16% nonresponse and another 23% truncated) presents a variety of challenges. Selection of models and predictor variables for imputation along with procedures for assessing the quality of imputations are presented. Not having key predictor variables (e.g., genetic indicators) in the imputation model due to not knowing of associations is a definite concern. Relevance of this research for large scale surveys over time is discussed.

Patterns of Landline & Cell Phone Usage & Implications

The Role of Landline and Cell Phone Usage Patterns in Nonresponse Error Potential Among Young Adults in Telephone Surveys

Douglas B. Currivan, RTI International (dcurrivan@rti.org); Burton Levine, RTI International (blevine@rti.org); Niki Mayo, RTI International (nmayo@rti.org)

A current challenge in conducting telephone surveys of adults is obtaining satisfactory representation among young adults. Surveys using traditional random-digit dial (RDD) sample frames of landline numbers exclude the 33 percent of adults age 18 to 24 who live in households without landline phone service. In addition to this well-documented coverage issue, a further problem is the potential difficulty in contacting and interviewing an additional set of about 19 percent of young adults who live in households with a landline phone, but primarily use a cell phone. RDD telephone surveys may further under represent the full population of young adults by excluding those who primarily rely on cell phones and are difficult to reach by landline phones. Young adults who primarily use cell phones may have important similarities to those who only have wireless phone service in terms of key survey measures.

To better understand the role of phone usage patterns in the potential for nonresponse error among young adults, this paper investigates the implications of patterns of landline and cell phone use among young adults for key survey outcomes. The data we use are drawn from a statewide RDD-based survey on health behaviors that targeted young adults age 18 to 24 and included both landline and cell phone numbers. Based on respondents' answers to questions on the presence, number, and sharing of cell phones in their household, we coded each respondent into one of four categories: landline only, primarily landline, primarily cell phone, or cell phone only. We then examined patterns in both demographic characteristics and health indicators across these four categories. This analysis provided data on how young adults' phone usage could potentially contribute to nonresponse error, beyond coverage error. We discuss the implications of these findings for the representation of young adults in RDD surveys on health behaviors.

Assessment of Bias in the National Immunization Survey–Teen: An Updated Benchmark to the National Health Interview Survey

Kennon R. Copeland, NORC at the University of Chicago (copeland-kennon@norc.org); Christina Dorell, Centers for Disease Control and Prevention (cdorell@cdc.gov); Meena Khare, Centers for Disease Control and Prevention (mkhare@cdc.gov); Abera Wouhib, Centers for Disease Control and Prevention (awouhib@cdc.gov); Elizabeth A. Ormson, NORC at the University of Chicago (ormson-elizabeth@norc.org)

Random-digit-dialing (RDD) surveys are subject to bias due to noncoverage of households with no access to landline telephones – both households with access only to wireless telephones (“wireless-only households”) and households with no telephone service (“nontelephone households”). The proportion of

wireless-only households has been increasing over time and is currently estimated at about 9 percent among households with teenagers aged 13-17 years while the proportion of nontelephone households has been relatively stable at roughly two percent. Similarly, potential for nonresponse bias in survey estimates increases as response rates decrease, but the actual level of bias must be evaluated for each survey. This paper evaluates potential noncoverage and nonresponse bias in estimates from the National Immunization Survey–Teen (NIS–Teen), an RDD-based telephone survey, based on comparisons with the National Health Interview Survey (NHIS), an address-based survey administered in person.

An initial comparison of the NIS–Teen and NHIS conducted using data from fourth quarter 2007 suggested that the respondents to the two surveys did not differ significantly overall. This updated analysis will use data from all four quarters of 2008 for both surveys, allowing a larger sample size and more robust analysis.

The 2008 NIS–Teen survey interviewed 30,725 households with adolescents aged 13-17 years. The survey collected information on doctor's visits, general health information, insurance status, and demographics. With parental consent, a mail survey was sent to the children's immunization providers to collect vaccination histories. Adequate vaccination histories were obtained from providers for 58% of adolescents with completed household interviews.

We compare socio-demographic distributions (e.g., family poverty level, maternal education, and race/ethnicity) and other health-related characteristics (presence of asthma, history of chickenpox, and health insurance) including parental report of influenza vaccination receipt in the past 12 months from the two surveys.

Landline and Cell Phone Usage Patterns in a Large Urban Setting: Results From The 2008 New York City Community Health Survey

Stephen Immerwahr, Survey Unit, Bureau of Epidemiology Services (simmerwa@health.nyc.gov); Michael P Battaglia, Abt Associates Inc. (mike_battaglia@abtassoc.com); Donna Eisenhower, NYC DOHMH Epidemiology Services (deisenho@health.nyc.gov); Michael Sanderson, NYC DOHMH Epidemiology Services (msander1@health.nyc.gov); Andy Weiss, Abt SRBI (A.WEISS@SRBI.COM); Bonnie Kerker, NYC DOHMH Epidemiology Services (bkerker@health.nyc.gov)

Little research has been presented on cell phone surveys conducted in large urban cities. New York City provides a unique opportunity as the largest and most diverse city in the country. The 2008 New York City Community Health Survey of adults used a dual frame sample design. We completed 7,554 interviews with adults in the landline RDD sample. In the cellular sample we screened 3,368 individuals.

The dual frame design allows us to identify and compare the demographic and health characteristics of four key telephone usage groups across the two samples – a) landline sample: landline only, dual service-cell mostly, dual service-not cell mostly; b) cellular sample: cell only, dual service-cell mostly, dual service-not cell mostly. We will focus primarily on differences between dual service-cell mostly adults in the landline and cellular samples, because most dual frame estimation schemes assume that these two sample groups are equivalent.

We will then report on telephone usage patterns in New York City based on a rich set of questions in the landline questionnaire and the cellular questionnaire related to the prevalence of: 1) cell phones used by persons under 18 years of age, 2) cell phone sharing among adults, 3) having more than one personal-use cell phone telephone numbers, 4) unlimited cell phone calling plans, 5) outside the home location

during the conduct of the cell phone interview, and 6) discontinuing landline telephone service in the next 12 months.

Implications for dual frame surveys conducted in large urban environments are discussed. Comparison with patterns found in national studies are compared.

List Assisted RDD Telephone Surveys in Rural Areas: Central Texas

Debbie L. McMahon, Baylor University (debbie_mcmahon@baylor.edu); Wesley M. Hinze, Baylor University (wes_hinze@baylor.edu)

Rural populations differ from urban populations in ways that create unique sampling issues when conducting telephone survey research. The Center for Community Research and Development (Baylor University) sampled 1,067 adult residents of five rural counties in Central Texas in the summer of 2009. Using a combination of prescreened list assisted telephone numbers purchased from a vendor and randomly generated telephone numbers created in-house, we carried out an experiment to determine the feasibility of using prescreened numbers to boost the contact rate in future telephone surveys of surrounding rural areas. Due to greater residential stability in rural areas, we hypothesized that the prescreened telephone numbers would produce a more cost-effective sample than the randomly generated numbers. Although prescreened number had better contact rates, the number of business, fax and nonworking numbers from the prescreened sample called into question the cost benefit of purchasing numbers for this rural survey. This suggests that surveys of rural households using list-assisted RDD samples may be less cost-effective than those that use pure random digit dial methodologies.

Utilizing Respondent Records

Respondent Record Use in the U.S. Consumer Expenditure Survey

Jennifer Edgar, Bureau of Labor Statistics (edgar_J@Bls.gov)

The US Consumer Expenditure Quarterly Interview Survey (CEQ) is a national household panel survey that asks respondents to report their household's expenses for a three month period. The respondent's recall task is very challenging, with respondents asked to report about an extensive variety of items, ranging from major appliances to clothing. It seems unlikely that, without the aid of records, respondents would be able to recall all expenditures, or report those they do recall very accurately. Ideally, respondents would use receipts and other financial records to complete the interview, though this would likely lengthen an already long interview, and may add considerable burden to the respondent.

This paper summarizes the CEQ procedures used to encourage respondent record use, and examines paradata to gain insight into the extent that respondents use records and the type of records used. We will also look at the impact of record use on interview length, data quality and attrition in later waves of the interview, and explore relationships between respondent characteristics and record usage to determine if there are categories of respondents more or less likely to use records.

The Use of Utility Bills as an Information Retrieval Aid in the American Housing Survey

George R Carter, U.S. Census Bureau (george.r.carter.iii@census.gov)

Utility costs are an important component of housing costs. The American Housing Survey (AHS) is a key source of housing cost and utility data for the United States. The AHS is collected by the U.S. Census Bureau for the Department of Housing Urban Development (HUD) and is used by HUD to track changes

in housing over time and to develop housing policy. Among the utility data the AHS collects are data on electric, gas, and oil bills, as well as bills for other fuels, trash collection, and water. To increase the accuracy of utility data reported, respondents to the AHS are encouraged to use actual bills to aid in information retrieval to the extent they are available. If respondents use electric or gas and are billed separately for these utilities, they are asked (1) if they have actual bills and (2) the amounts of those bills for January, April, August, and December. To avoid misreporting due to recall problems, respondents who have no bills are asked to report only the last bill they received. Using data from the 2007 AHS on occupied housing units, this presentation presents research on respondents' use of bills in answering questions about electric and gas utilities. Demographic, economic, and regional variations in bill usage are examined. Preliminary research suggests that a majority of respondents do not have bills available to aid in answering questions about specific months and that householders who are owners, older, and have higher levels of education are more likely to have complete billing records. Methods of encouraging record use in the AHS and of assessing the validity of responses using records will be explored and implications of record use on data quality will be discussed.

For Clueless Consumers, Can Records Improve Data Quality?

Eileen M O'Brien, Energy Information Administration (eileen.obrien@eia.doe.gov)

Estimating household expenses usually requires that respondents perform complex cognitive tasks associated with records use, recall or both. This paper describes a more controlled and limited use of household records by respondents given features of the data requested--annual energy consumption and expenditures. Energy bills include special charges, some which are out of scope. Respondents do not understand these charges or consumption metrics well. The reference period is long (12 to 16 months) relative to household recordkeeping practices and the data collection schedule. For household budgeting purposes, monthly bills may be 'levelized' over a year by the utility and not relate to real consumption, which depends a lot on weather, or prices, which depend a lot on deregulated markets. Energy bills are simply too difficult for respondents to use in a timely, reliable and unburdensome way. The alternative approach, then, has been to have respondents provide a minimal amount of information to conduct a follow-on survey of their energy suppliers: a monthly utility bill with an account number for each supplier. One feature of this two-phase design has ensured the success of this type of records use over thirty years: participation by householders is voluntary, but participation by energy suppliers is not. Nevertheless, more accommodation and new processes have been required to maintain data quality because of changes in the following: the survey taking climate, household composition, federal confidentiality protections, state laws, company practices, and how suppliers maintain and report data. This two phase design preserves the use of important household records by shifting response burden to the record experts: the companies that produce them. This paper offers considerations for the field as we use more administrative records in lieu of household-reported data to manage survey error, burden and costs.

The Burden of Proof: Panel Attrition and Record Usage on the MCBS

Brad Edwards, Westat (BradEdwards@westat.com); Ryan A Hubbard, Westat (RyanHubbard@westat.com)

The paper explores the impact of various types of respondent burden on panel attrition in the Medicare Current Beneficiary Survey (MCBS). The study is designed to provide a descriptive account of burden-based attrition on the MCBS through a comparative analysis of fee-for-service Medicare respondents who produce statements for use in the interview and respondents who do not produce statements. Specifically, we focus on the relative effect of record usage on non-death attrition in subsequent waves.

Record usage includes the collection and abstraction of information from insurance statements, when available, in an effort to accurately record health care utilization and costs. Recordkeeping is a major part of respondent burden on the MCBS, but other factors are also at play in panel attrition. Respondent characteristics such as demographics and self-reported mental and physical health; interview characteristics such as interviewer experience, the introduction of a new interviewer in subsequent waves, changes in respondent, and the complexity or length of the interview; and study characteristics such as topic salience, sponsorship, and perceived burden have all been shown to impact non-response. Through logistic regression and survival modeling of study data, sample frame information, and paradata, this paper details the relative effect of record usage and other factors on panel attrition.

***Is it Worth the Effort? Respondent Record Use on the British Household Panel Study
(For the Respondent Record Use Panel)***

Nick Moon, GfK NOP Social Research (nick.moon@gfk.com); Heather Laurie, ISER, University of Essex (laurh@essex.ac.uk)

The British Household Panel Study is a major UK-wide study that ran for fifteen annual waves before being merged into the new, and larger, UK Longitudinal Household Survey in 2009. The BHPS involves around 15,000 interviews in around 9,000 households across the UK, with attempts made to interview every person aged 16 and over in each interviewed household.

One of the main topics covered by the multi-topic questionnaire is employment, and in particular income from it. There are detailed questions on gross and net pay on the last occasion respondents were paid and normal pay. Respondents are asked to check their most recent payslip to ensure the information is as accurate as possible, and typically around a third do so each wave.

In this paper we estimate the improvement the use of payslips produces in data quality, and also the wider issues such as whether certain types of respondents are more likely to consult payslips, whether there is any correlation between consultation of payslips and long-term co-operation in the survey, and any differences in the length of the employment section of the questionnaire when payslips are and are not used.

The paper also reports on a small-scale survey of interviewer opinion on the process, covering any impressions of concern among respondents about being asked to consult payslips, whether those who consulted them had them to hand or had to go and find them, and comparing BHPS with other surveys where respondents have been asked to undertake similar tasks